Metacommunity dynamics and the spurious detection of species associations in co-occurrence analyses: patch disturbance matters

Vincent Calcagno [1,†], **Nik J. Cunniffe** [2,†,*], **Frédéric M. Hamelin** [3,†]

[1] Université Côte d'Azur, INRAE, CNRS, Institut Sophia Agrobiotech, Sophia-Antipolis, France;
[2] Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom;
[3] IGEPP, INRAE, Institut Agro, Univ Rennes, Rennes, France
† Authors are listed in alphabetical order and share equal co-authorship.
* Corresponding author: `njc1001@cam.ac.uk`

Abstract Many methods attempt to detect species associations from co-occurrence patterns. Such associations are then typically used to infer inter-specific interactions. However, correlation is not equivalent to interaction. Habitat heterogeneity and out-of-equilibrium colonization histories are acknowledged to cause species associations even when inter-specific interactions are absent. Here we show how classical metacommunity dynamics, within a homogeneous habitat at equilibrium, can also lead to statistical associations. This occurs even when species do not interact. All that is required is patch disturbance (i.e. simultaneous extinction of several species in a patch) a common phenomenon in a wide range of real systems. We compare direct tests of pairwise independence, matrix permutation approaches and joint species distribution modelling. We use mathematical analysis and example simulations to show that patch disturbance leads all these methods to produce characteristic signatures of spurious association from "null" co-occurrence matrices. Including patch age (i.e. the time since the last patch disturbance event) as a covariate is necessary to resolve this artefact. However, this would require data that very often are not available in practice for these types of analyses. We contend that patch disturbance is a key (but hitherto overlooked) factor which must be accounted for when analysing species co-occurrence.

Keywords: extinction, disturbance, permutations, metapopulation, independence, interactions, Sim2, Sim9, JSDM, Hmsc

1 Introduction

Community ecology has long been concerned with inferring interactions between species (e.g. competition or facilitation) from co-occurrence data (Forbes, 1907; Cohen, 1970; Diamond, 1975; Caswell, 1976; Connor and Simberloff, 1979). Recently, metagenomics approaches have renewed this use of co-occurrence patterns as a means to infer species interaction networks, in environments ranging from soils to the human microbiome (Barberán et al., 2012; Faust and Raes, 2012). If species do not interact, intuition suggests the proportion of patches (or hosts) where species co-occur should be the product of the proportions of patches occupied by each species. Testing for pairwise independence, i.e. looking for statistical patterns of association, consequently underpins most methods for inferring species interactions from presence-absence data (Gotelli, 2000; Gotelli and Ulrich, 2012; Ovaskainen and Abrego, 2020). Positive associations (excess of co-occurrences) may indicate facilitation whereas negative associations (deficit of co-occurrences) may indicate competition.

However, it is well known that several pitfalls can compromise this approach, since correlation is not equivalent to interaction (Blanchet et al., 2020; Molina and Stone, 2020). As a simple example, patch heterogeneity and differential habitat preferences among species can cause species associations, even in the absence of any interaction between the pairs of species involved. Similarly, if some patch types host more species than others overall, because of their larger size or more profitable conditions, the occurrence of a species in a patch increases the odds of occurrence of other species, simply because it makes it more likely the patch is of higher quality. This would generate positive association signals if not controlled for. Out-of-equilibrium dynamics, i.e. species expanding through space, concomitantly or differentially, are also known to cause similar patterns. Different statistical methods have been proposed to take into account such patch heterogeneity. However, it is commonly expected than in the absence of any of the above elements, or after having appropriately controlled for them, classical extinction/recolonization dynamics (metacommunity dynamics; Leibold et al., 2004) would not introduce particular species associations, unless ecological interactions are actually occurring (Opedal et al., 2020).

Here we show that a specific feature of metacommunity dynamics has been overlooked so far: patch disturbance, i.e. the stochastic occurrence of patch extinctions (Hastings, 1980, 2003). Patch disturbance events denote extreme but possibly frequent events, such as fires, droughts, floods and others, causing the local extinction of all, or perhaps some subset of all, species from a patch (Sousa, 1984; Leibold et al., 2004). This can also correspond to the actual destruction and disappearance of the patch, for instance in ecosystems managed by humans, such as agricultural fields or forest stands (Ovaskainen et al., 2017), or in systems (parasite or microbiome data) for which "patches" are actually hosts with a finite lifespan (Hamelin et al., 2019; Friedman and Alm, 2012). Patch disturbances are ubiquitous in natural and anthropogenic systems (Pickett and White, 2013; Jentsch and White, 2019). Although commonly incorporated in metapopulation and metacommunity frameworks (e.g. Levin and Paine, 1974; Hastings, 1980, 2003; Leibold et al., 2004; Calcagno et al., 2011), patch disturbances are sometimes omitted in favour of species-specific extinctions only, as a seemingly benine mathematical simplification (e.g. Slatkin, 1974; Thompson et al., 2020).

We introduce a general metacommunity model of patch dynamics, that allows for patch disturbance to occur at some, potentially time-varying, rate, and in which patches can have a finite lifespan. This model generalizes many existing particular models. The model is “null” in the sense that species are not interacting, the habitat is completely homogeneous, and all species disperse uniformly with no patch preferences. We analyze the equilibrium occupancies and co-occurrence patterns, from which we generate sample co-occurrence matrices. We then test different classical approaches to detect species associations on these “null” synthetic data. A first approach (e.g. Gotelli, 2000, Sim2) bluntly tests for species pairwise independence and does not correct for patch differences, while a second approach (e.g. Gotelli, 2000, Sim9) controls for patch differences in overall richness. Both are shown to yield characteristic patterns of spurious species association, caused by patch disturbance events and species differences in extinction/colonization dynamics. We explain these results mathematically by analyzing the probabilities of species co-occurrence, and the action of the statistical methods. We show that a quantity we term species “fastness”, that quantifies how quickly species occurrence probability

increases after disturbance, is key to predict the patterns of spurious associations detected. We predict and show that a third approach, using joint species distribution modelling with patch age (i.e. time since last patch disturbance, e.g. Ovaskainen et al., 2017) as a covariate, is required to identify the correct conclusion, i.e. lack of spurious association signals. Since patch age would often not be known, we explore whether species richness may suffice as a proxy for patch age. We find it does not, suggesting explicit data on patch age might generally be required to appropriately control for metacommunity dynamics effects in co-occurrence analyses.

2 Model & Methods

2.1 General metacommunity model

We consider a metacommunity model describing the extinction-recolonization dynamics of s non-interacting species over a large number of identical patches (Caswell, 1976). Each species has its own colonization rate c_i per occupied patch, as well as a constant rate of immigration from outside the metacommunity, m_i (Hastings, 1987). Any species can go extinct in a patch, at some rate e_i. In addition, patches can undergo catastrophic disturbances, after which all species in the patch go extinct, irrespective of species composition (Hastings, 1980; Leibold et al., 2004; Calcagno et al., 2011). Such catastrophes occur at rate μ_x, where x is the "age" of a patch, i.e. the time since its appearance or since it experienced the last catastrophic disturbance event (Levin and Paine, 1974; Hastings, 2003). Finally, there may exist a maximal patch age X after which a catastrophic disturbance systematically occurs (Olivieri et al., 1995). The model is illustrated in Figure 1.

Let $p_{i,x,t}$ be the fraction of patches that have age x and are occupied by species i at time t (Hastings, 1991). The fraction of patches occupied by species i is $p_{i,\bullet,t}$, and the fraction of patches with age x at time t is $p_{\bullet,x,t}$. The fraction of patches with age x that are not yet occupied by species i is therefore $p_{\bullet,x,t} - p_{i,x,t}$ (see Figure 1A). We refer to Supporting Information A (section A.1.1) for more technical definitions. The general metacommunity model can be written as, for all $i = 1, 2, \ldots, s$,

$$\frac{\partial p_{i,x,t}}{\partial x} + \frac{\partial p_{i,x,t}}{\partial t} = -(\mu_x + e_i)\, p_{i,x,t} + (c_i p_{i,\bullet,t} + m_i)(p_{\bullet,x,t} - p_{i,x,t}) \,. \tag{1}$$

This framework generalizes classical metacommunity models (e.g. Cohen, 1970; Slatkin, 1974; Hastings, 1980; Hanski, 1982; Hastings, 1987; Tilman, 1994; Tilman et al., 1994; Hanski and Gyllenberg, 1997; Bell, 2001), deriving from MacArthur and Wilson (1963)'s "mainland-island" model ($c_i = 0$), and from Levins (1969)'s metapopulation model ($m_i = 0$). Such metacommunity models, when they consider patch catastrophic extinctions at all, assume a constant disturbance rate $\mu_x = \mu$, but this needs not be the case. The rate of disturbance might be increasing with patch age, producing an accelerating failure time model, for instance if a patch is a host that ages and suffers higher mortality with aging; this is termed a Type 1 survivorship curve in Begon and Townsend (2020). Conversely, if patches have some variability in their risk of disturbance, younger patches would go extinct at a relatively high rate, whereas older patches are relatively more disturbance resistant, causing a decline of μ_x with patch age (Type 3 survivorship curve). Another common situation is when patches have a finite lifetime and systematically get destroyed after some prescribed duration X (Pickett and White, 2013), as may happen for instance in agricultural settings where crops are harvested after some finite time. Our framework encompasses all such situations. As special cases, the classical "mainland-island" and "Levins" models correspond to $c_i = 0$ and $m_i = 0$, respectively (Gotelli, 1991; Hanski and Gyllenberg, 1993), together with $X \to \infty$ and $\mu_x = \mu$, a constant (see Supporting Information A, section A2).

2.2 Steady-state occupancies and co-occurrence patterns

At steady-state, we can drop the t subscripts. The overall occupancy of species i, $p_{i,\bullet}$, can be derived explicitly in the special cases of the mainland-island or Levins models, for which simple expressions exist (Supporting Information A, section A2.1). For instance, $p_{i,\bullet} = 1 - (e_i + \mu)/c_i$ in the Levins model. In general, no such solution exists when the rate of patch disturbance depends on age, but $p_{i,\bullet}$ can be computed numerically, e.g. by using a recursive algorithm outlined in Supporting Information A, section A1.4.

It would be possible to derive a separate equation for the dynamics of every possible patch state, i.e. the fraction of patches that are occupied by any given combination of species, and to solve these (as done in Supporting Information A,

section A2.2, for two species). However, this is computationally inefficient. To obtain random co-occurrence matrices drawn from the above model, we rather use a Monte-Carlo procedure that generates the desired number of patches and draws their species composition. We provide a comprehensive set of R functions that make it easy to compute the steady-state occupancy of each species and generate random species co-occurrence matrices from the general model 1, for any number of species, patches, and any shape of the disturbance function μ_x (Supporting Information B). We will use this method to generate null co-occurrence matrices that involve no species interactions, just metacommunity dynamics (Figure 1).

2.3 Introducing relative distribution profiles and 'fastness'

As illustrated in Figure 2A, at steady state there exists a stable distribution of patch ages, and species occupancies vary as a function of patch age. We here derive a quantity $\pi_{i/x}$ representing how species i is distributed over patch ages. We term this quantity the relative distribution profile of a species (Figure 2B). Values smaller (larger) than one means the species i is rarer (more frequent) in patches of age x, relative to its overall occupancy $p_{i,\bullet}$.

We first define the fraction of patches of age x that are occupied by species i as $p_{i|x} = p_{i,x}/p_{\bullet,x}$, for which we obtain an explicit expression from model 1 at steady-state (Supporting Information A, section A1.3). The relative distribution profile of species i is then derived as the fraction of patches of age x that are occupied by species i, relative to its overall probability of occupancy $p_{i,\bullet}$:

$$\pi_{i/x} = \frac{p_{i|x}}{p_{i,\bullet}} = \frac{1}{p_{i,\bullet}} \frac{c_i p_{i,\bullet} + m_i}{c_i p_{i,\bullet} + m_i + e_i} \left(1 - \exp\left(-(c_i p_{i,\bullet} + m_i + e_i)x\right)\right) . \tag{2}$$

As shown in Figure 2B, the mean value of the profile, for any species, is one by definition (Supporting Information A, section A1.5).

To facilitate interpretation, we will discriminate species based on their "fastness": fast species have very flat relative distribution profiles that quickly jump from zero and saturate at a value close to one (Figure 2B). Slow species, on the contrary, have relative distribution profiles that slowly increase from zero and reach much larger values as the patch age gets large.

Formally, we will quantify the fastness of a relative distribution profile as its variance over patch ages, $\mathrm{V\alpha r}(\pi_{i/x})$. A maximally fast species has $\pi_{i/x} = 1$ for all x, i.e. zero variance (Figure 2). Conversely, a very slow species has a relative distribution that gradually climbs to very large values, implying a huge variance. It can be shown that the variance of a species is also strictly decreasing with the initial slope of its relative distribution profile (Supporting Information A, section A1.7). The variance of a species profile is thus a metric of "slowness", and its inverse is a metric of fastness. Importantly, the fastness of a species is not directly related to its overall occupancy in the metacommunity.

2.4 Testing for species associations in co-occurrence matrices

The probability of co-occurrence of two species i and j, i.e. the overall fraction of patches in which the two species are found, is denoted $q_{i,j,\bullet}$ (at steady-state). This is often used as a metric of species association, that can be estimated from species co-occurrence matrices, and tested for deviations from "null" expectations. A simple null model is that of species independence: if species are not interacting, they should be distributed independently across patches, and thus the value of $q_{i,j,\bullet}$ is compared to the null value $p_{i,\bullet}p_{j,\bullet}$ using a standard test for contingency tables (Veech, 2013).

This prediction is straightforward and intuitive, but is known to suffer from several limitations. To circumvent these, more sophisticated null models are often preferred. A widespread approach consists in using permutation schemes intended to break species associations in co-occurrence matrices while retaining important species and patch differences (Gotelli, 2000). In this context, the C-score (Stone and Roberts, 1990) is the metric most commonly used to quantify the tendency of species to be segregated or aggregated in co-occurrence matrices. The partial C-score between two species, $C_{i,j}$, can be expressed as

$$C_{i,j} = N^2(p_{i,\bullet} - q_{i,j,\bullet})(p_{j,\bullet} - q_{i,j,\bullet}) \,,$$

with N the total number of sites (patches) in the matrix.

Values of $C_{i,j}$ larger than expected indicate segregation (e.g. competition) between the two species, whereas values smaller than expected indicate preferen-

tial association (e.g. facilitative interactions). Several null models have been proposed to determine the expected $C_{i,j}$ value and test for deviations from it. The most classical methods are the so-called fixed-equiprobable and fixed-fixed permutation schemes, that reshuffle matrix values while keeping the row sums fixed, or both the row and column sums fixed, respectively (Gotelli and Ulrich, 2012; Münkemüller et al., 2020).

More recently, flexible methods relying on the statistical modelling of species occurrences based on various fixed or random predictors have gained a lot of popularity. These allow to infer species associations from their residual covariances. For instance, hierarchical models of species composition (Hmsc) are increasingly used (Tikhonov et al., 2020; Ovaskainen and Abrego, 2020).

In the following, we will analyze how null co-occurrence matrices generated from our general metacommunity model behave, in terms of producing significant species associations, when subjected to the different approaches described above (Figure 1). We will derive mathematical predictions and illustrate them numerically using a synthetic community of 31 species, chosen to harbour variation in their occupancies and extinction/colonization parameters (Figure 3A; see Supporting Information B for details). These are not real species, but the range of variation they present in their colonization rates is similar to that found in natural metacommunities, for instance in Caribbean pond snails (Dubart et al., 2019) or grassland plant communities (Tilman, 1994).

As our illustrative parameter set, we assumed no external immigration ($m_i = 0$), with disturbance rate $\mu(x) = 0.2$ per unit time (a constant) and $X = 20$ time units (maximum lifespan of a patch). This rate of disturbance is intermediate relative to species-specific extinction/colonization parameters (average extinction rate 0.5 per unit time, average colonization rate 1.5). All numerical simulations and figures can be reproduced using the R markdown provided (Supporting Information B).

3 Results

3.1 Patch disturbance generates spurious positive associations if patch age is not accounted for

In the absence of any interaction, species are independently distributed among patches, within any particular age class (Supporting Information A, section A3). This implies

$$q_{i,j,\bullet} = \int_0^X p_{i|x} p_{j|x} p_{\bullet,x} \mathrm{d}x \,, \tag{3}$$

which can be equivalently expressed as:

$$q_{i,j,\bullet} = p_{i,\bullet} p_{j,\bullet} \int_0^X p_{\bullet,x} \pi_{i/x} \pi_{j/x} \mathrm{d}x = p_{i,\bullet} p_{j,\bullet} (1 + \mathrm{Cov}(\pi_{i/x}, \pi_{j/x}))\,, \tag{4}$$

given that the mean of the profile is one (Supporting Information A, section A1.5). Since $\pi_{i/x}$ and $\pi_{j/x}$ are increasing functions of x (Eq. 2), from Harris' inequality we have $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) \geq 0$ (Supporting Information A, section A1.5). From (4) we therefore have

$$q_{i,j,\bullet} \geq p_{i,\bullet} p_{j,\bullet}\,. \tag{5}$$

Equality occurs only if at least one of the two species is distributed uniformly over all patch ages, i.e. $\pi_{k/x} = 1$ for all x, so that $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) = 0$. From Eq. 2, this is not the case in general. This only occurs if there is no patch disturbance at all ($\mu_x = 0$ and $X \to \infty$), in which case all patches are effectively infinitely old, so that $\pi_{i/x}$ and $\pi_{j/x}$ both tend to one, and $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x})$ tends to zero. But in all other cases, species are not distributed independently: they are positively associated, i.e. they co-occur more often than expected from the "null" product of their respective overall occupancies (see Eq. 5).

As a consequence, methods based on the test of pairwise independence of species (e.g. Veech, 2013) would consistently detect spurious positive associations (aggregation) among species, even in the absence of any interaction between species. Specifically, positive associations will be strongest, and thus more likely to be statistically detected, for species pairs with a large covariance of their relative profiles (Eq. 4), i.e. for pairs of species that are both slow. In contrast, species pairs in-

cluding at least one fast species are more likely to be declared independent. As a result, a block-like pattern of significant positive interactions will result, if species are ordered according to their fastness. These predictions are confirmed numerically, as illustrated in Figure 3, using Fisher's test to assess pairwise independence.

The same conclusion holds for other methods that do not control for patch age, for example the "fixed equiprobable" permutation algorithm, also known as "Sim2" (Gotelli, 2000). To show this, we take the commonly-used partial C-score $C_{i,j}$ as a metric for species association. We will denote with an asterisk the value of quantities after the permutation algorithm has been applied to a co-occurrence matrix.

The fixed-equiprobable matrix permutation algorithm preserves the number of occurrences per row (i.e. overall occupancies of species) but reshuffles species across all patches equiprobably. Therefore, after permutations, species overall occupancies are unchanged ($p^*_{i,\bullet} = p_{i,\bullet}$ for all i), but all species become equiprobable over the different age classes, on average. Therefore, their relative distribution profiles become flat: $\pi^*_{k/x} = 1$ for all x. From Eq. 4 it follows

$$q^*_{i,j,\bullet} = p^*_{i,\bullet} p^*_{j,\bullet} = p_{i,\bullet} p_{j,\bullet} \,. \tag{6}$$

This indicates that the algorithm effectively breaks the positive association generated by patch disturbance. From Eq. 5, this implies $q^*_{i,j,\bullet} \leq q_{i,j,\bullet}$. Permuted matrices will on average have a deficit of co-occurrences, for all species pairs, compared to the original matrix. Since $p_{i,\bullet}$ and $p_{j,\bullet}$ are unaffected by permutation, we can directly conclude that $C^*_{i,j} > C_{i,j}$, as soon as patch disturbances occur. Using the fixed-equiprobable permutation algorithm will therefore consistently yield spurious facilitation signals between all pairs of species, exactly as direct tests for species independence do (see Fig. C1A in Supporting Information C).

Similarly, ordination methods and joint species distribution models that do not include patch-age as an explicit covariate will return spurious positive residual associations among most species pairs, except those including at least one very fast species. This is illustrated numerically in Fig. C1C in Supporting Information C.

3.2 Patch disturbance yields a characteristic signature of spurious associations even if patch richness is controlled for

We now turn to methods that can control for systematic differences among patches, in this case emerging from heterogeneity in the time since last disturbance (patch age). One such method is the classical fixed-fixed permutation algorithm, also known as Sim9 or swap (Gotelli, 2000). By fixing the row totals, this method effectively maintains the number of species per patch unchanged, and can thus deal with the positive association of species reported above.

Similar to the previous permutation algorithm (Sim2), preservation of row totals implies $p^*_{i,\bullet} = p_{i,\bullet}$. Therefore, we only need to study changes in $q_{i,j,\bullet}$ in order to understand changes in C-scores. Preservation of column totals further means that the number of occurrences will remain the same in every age class, which can be stated as $\sum_{k=1}^{s} p^*_{k,x} = \sum_{k=1}^{s} p_{k,x}$ for all x. Species are otherwise reshuffled indifferently across all patch ages. Let w_i be the relative occupancy of species i in the matrix, that remains unchanged after permutations:

$$w_i = \frac{p_{i,\bullet}}{\sum_{k=1}^{s} p_{k,\bullet}} \,.$$

A fixed-fixed permutation algorithm yields:

$$p^*_{i,x} = \left(\sum_{k=1}^{s} p_{k,x} \right) w_i \,,$$

leading to, for all $i = 1, 2, \ldots, s$,

$$\pi^*_{i/x} = \frac{p^*_{i,x}}{p_{\bullet,x} p_{i,\bullet}} = \frac{\sum_{k=1}^{s} p_{k,x}}{p_{\bullet,x} \sum_{k=1}^{s} p_{k,\bullet}} = \frac{\sum_{k=1}^{s} p_{k,\bullet} \pi_{k/x}}{\sum_{k=1}^{s} p_{k,\bullet}} = \sum_{k=1}^{s} w_k \pi_{k/x} = \hat{\pi}_x \,.$$

This indicates that species relative distribution profiles are replaced with their weighted-average over all species, with weights equal to species relative overall occupancies. Let us denote this weighted-average by $\hat{\pi}_x$.

We thus obtain:

$$q^*_{i,j,\bullet} = p_{i,\bullet} p_{j,\bullet} \int_0^X p_{\bullet,x} \pi^*_{i/x} \pi^*_{j/x} \mathrm{d}x = p_{i,\bullet} p_{j,\bullet} \int_0^X p_{\bullet,x} \hat{\pi}_x^2 \mathrm{d}x = p_{i,\bullet} p_{j,\bullet} (1 + \mathrm{V\alpha r}(\hat{\pi})) \,. \qquad (7)$$

Comparing with Eq. 4, we see that the term correcting for non-independence, $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x})$, is here replaced with $\mathrm{Var}(\hat{\pi})$. The difference $q_{i,j,\bullet} - q^*_{i,j,\bullet}$ has the same sign as $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) - \mathrm{Var}(\hat{\pi})$.

The permutation algorithm is thus biased, except if $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) = \mathrm{Var}(\hat{\pi})$ for all species pairs. This would occur if all species had identical relative distribution profiles, i.e. when all species are "similar", in the sense $p_{i,x} = \kappa p_{j,x}$ for all x, with some positive κ. This trivially occurs if there is no patch disturbance at all, or if all species have identical extinction and colonization parameters. Beyond that, species may differ in their parameters and overall occupancies and still be similar in that sense, but this require quite constrained parametric conditions (Supporting Information A, section A2.3).

Except for these restricted scenarios, the permutation algorithm will yield spurious associations in both directions. Species pairs whose relative distribution profiles have covariance smaller than the variance of the average community profile ($\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) < \mathrm{Var}(\hat{\pi})$) will appear to have a spurious negative association. Conversely, species pairs for which $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) > \mathrm{Var}(\hat{\pi})$ will appear to have a spurious positive association. Only species pairs for which $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) = \mathrm{Var}(\hat{\pi})$ would appear as independently distributed.

In practical terms, species pairs containing at least one species that is 'faster' than average will have low relative covariance, and thus will be declared negatively associated. Indeed if the fast species (say i) is fast enough (regardless of how slow the other is), then $\pi_{1/x} \approx 1$ for all x thus and $\mathrm{Cov}(\pi_{i/x}, \pi_{j/x}) \approx 0 = \mathrm{Var}(\pi_{i/x})$, which is less than the variance of the average. In contrast, species pairs that are both 'slower' than average will have large relative covariance, and will be declared positively associated. Pairs of species that are both close to average fastness may yield no particular signal.

If species are ordered according to fastness (i.e. variance), this produces a characteristic signature of spurious association signals. A numerical example is presented in Figure 4A, from the same co-occurrence matrix as in Figure 3. We observe spurious associations of both kinds, as expected, even though the total number of species pairs for which a spurious association is reported is slightly smaller than it was without controlling for patch richness (56% versus 73%).

Controlling for differences across patches is therefore insufficient to handle the metacommunity dynamical consequences of patch disturbance, even though it brings some improvement: first, correct results are obtained if all species are "similar" (i.e. have identical fastnesses), second, the overall number of spurious associations obtained is reduced. However, this generates spurious associations of both kinds (positive and negative), not only positive associations.

The same conclusions apply to other similar methods, for instance to joint species distribution models in which differences among patches are represented with a single latent factor (as hypothesized in Ovaskainen and Abrego, 2020, p. 335; see Supporting Information B for details). This is shown in Figure 4C. The signature of spurious signals is similar, with the difference that fast-fast species pairs (bottom left corner) are declared positively associated as slow-slow species pairs, producing a four-block pattern reminiscent of differential specialization on two habitats.

Throughout, we assumed all species in the metacommunity were affected by patch disturbances. In practice, in large sets of species, it may be that only a subset of species are susceptible to disturbance events, while others are not ("immune" species). Based on our theoretical results, this is straightforward to deal with: such immune species would behave as species with infinite fastness, regardless of their colonization/extinction rates (Supporting Information A, section A4). Therefore, any species pair including at least one immune species is expected to return no overall association (if using independence tests and related methods, as in Section 2.1; see Fig. C1B and C1D in Supporting Information C). If using a fixed-fixed permutation method, such species pairs are expected to generate spurious negative association signals, on top of the other ones. This prediction is illustrated in Figure 4B, in which six of the 31 species were rendered immune to patch disturbance. Similarly, if using a Hmsc approach, the presence of immune species, by inflating the heterogeneity among species, reinforces the pattern of spurious associations, and immune species tend to cluster with fast species. This is shown in Figure 4D.

3.3 Patch disturbance requires explicit modelling of patch-age

Efficiently dealing with patch disturbance in co-occurrence analyses therefore requires taking into account both patch age and species traits, in order to model ap-

propriately the differential effect that patch-age has on species with different fastnesses. In other words, there must be an interaction between the effect of patch-age and the identity of species.

This cannot be done with simple null model approaches, but can be in the context of joint species distribution models. An example is shown in Figure 5A. Using the same set of species and co-occurrence matrix as before, but using the actual patch-age as an environmental covariate in the Hmsc model, we were able to suppress the spurious association signals, as desired. By regressing the probability of occurrence on patch age, for every species individually, such a model appropriately captures both the overall increase in occupancy with patch age (Figure 5B) and the differential response of the different species, depending on their fastness (Figure 5C).

Note that this approach requires providing additional data on the age of patches, on top of the species co-occurrence matrix. A simpler alternative would be to use patch richness as a surrogate for patch age, considering the fair association that exists between the two, and the fact that the patch richness is directly encoded in the species co-occurrence matrix already (Figure 5B). However, using patch richness as a covariate instead of the actual patch-age fails to provide correct results (Figure 5D). Such a Hmsc model actually performs no better than if modelling patch richness as a single latent factor (Figure 4C), presumably because the correlation of patch-age and patch richness is insufficiently strong.

Our main conclusions are synthesized in Table 1.

4 Discussion

Metacommunity dynamics matter. We demonstrated here how simple metacommunity dynamics can cause the detection of spurious species associations, even in null communities with no interactions. This occurs as soon as patch disturbances exist, i.e. catastrophic events that cause all species (or at least a subset of them) to be locally wiped out. The latter is an arguably common element in natural and exploited ecosystems (Pickett and White, 2013), and one often, though not always, included in metapopulation and metacommunity models (Leibold et al., 2004).

It is classically assumed that in the absence of any ecological interactions, habi-

tat heterogeneity, or species differences in dispersal patterns, species would be distributed independently across patches. This “null” assumption underlies many interpretations of species co-occurrence patterns. Our analysis shows that this intuition is true for any given patch age (time since last disturbance), but it is not overall. Patch age acts as a hidden structuring variable, whose implications for co-occurrence patterns have been overlooked so far. Indeed, among the list of processes susceptible to cause departures from species independence, patch age does not normally appear (Gotelli, 2000; Ovaskainen and Abrego, 2020).

This result may seem counter-intuitive at first glance. Indeed, if species are initially independently distributed in a metacommunity, patch disturbances do not *per se* cause any association to appear: as they occur regardless of species composition, they preserve species independence, unlike interaction-driven extinctions would. Rather, it is the outcome of extinction/recolonization dynamics (Eq. 2) that inevitably causes, at steady-state, species to be more frequent in older patches. Since all species are more likely to occupy older patches, all appear positively associated overall.

The consequences of patch disturbance cannot simply be dealt with by controlling for patch differences in richness. If doing so, spurious associations, both negative and positive, also appear, as soon as species have different relative distribution profiles (fastnesses). These spurious associations have a different cause: they stem from the fact that slow and fast species are concentrated in different parts of the patch-age spectrum, what can be thought as a form of niche differentiation. This differential distribution w.r.t. to patch age causes spurious negative association signals to appear. But the signature we report is more subtle than that (see Figure 4). The fastest species (or species immune to patch disturbance, if any) also appear to be segregated from each other, as well as from all the other species. Reciprocally, the slowest species, being more concentrated than others in the older parts of the patch age spectrum, appear positively associated (and segregated from the others).

The generation of spurious associations (Type I error) is in itself problematic. But beyond that, in more realistic communities with actual species interactions, these spurious patterns would interfere with and confound the actual signals. For instance, actual signals of species interactions, e.g. an increase in species extinction rates

with species number, i.e. a “neutral” form of competition (Hastings, 1987; Bell, 2001), could be hidden in the presence of patch disturbance: the negative association patterns would be cancelled, at least among pairs of slow species, by the positive signal caused by metacommunity dynamics alone. In such situations, false negatives (loss of power) will result. Furthermore, the power of detection is not homogeneously distributed across all species pairs: signals of competition would be weakened for slow species pairs, but reinforced for fast species pairs. Therefore, a diffuse type of competitive interaction with no discernible pattern (all species competitively identical) could turn into a structured pattern, with non-zero nestedness.

A possible solution to these issues consists of using species distribution models that explicitly incorporate patch age as a covariate (Figure 5). However, this requires having independent data on patch ages, which is far from being universal in co-occurrence analyses. Whilst feasible for some types of systems (Ovaskainen et al., 2017), typically parasite or microbiome communities for which patches are hosts of known age, in many cases such data would be difficult or impossible to obtain. It may sometimes be possible to approximate patch-age from measurable environmental variables, when patch disturbance causes environmental modifications and patches follow a succession-like cycle after them. However, this is far from universal, and a sufficiently strong correlation would probably be required. Indeed, in our examples, using patch richness as a proxy of patch-age failed to remove the spurious associations, even though the two variables were reasonably well correlated (Figure 5). In general, effectively controlling for patch age would almost certainly require genuine temporal data on patch life-cycles (Ovaskainen et al., 2017, Fig. 5c). This underscores once more how inferring species interactions from snapshots of co-occurrence patterns can be as hazardous as it is appealing, and requires careful thinking about null expectations (Blanchet et al., 2020; Molina and Stone, 2020; Münkemüller et al., 2020). We join the recent burst of cautionary articles in highlighting these concerns, and in any case, we encourage investigators to evaluate whether detected species association patterns could match those reported here, and so could stem from pure metacommunity dynamical effects.

Acknowledgments The authors thank Maxime Dubart for guidance on using Hmsc and for sharing data on Caribbean snail metacommunity parameter estimates. Colin Guétemme and Héloïse Villesèche participated in preliminary research on the topic.

Data availability statement All the code used to produce the results in this article is available on GitHub at https://github.com/nikcunniffe/MetacommunityDynamics.

Conflicts of interest statement The authors declare no conflict of interest.

Author contributions Authors contributed similarly to the manuscript preparation. VC and FMH contributed most to the mathematical derivations, NJC contributed most to the stochastic simulations and R coding.

References

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal*, 6(2):343–351.

Begon, M. and Townsend, C. R. (2020). *Ecology: from individuals to ecosystems*. John Wiley & Sons.

Bell, G. (2001). Neutral macroecology. *Science*, 293(5539):2413–2418.

Blanchet, F. G., Cazelles, K., and Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*.

Calcagno, V., Massol, F., Mouquet, N., Jarne, P., and David, P. (2011). Constraints on food chain length arising from regional metacommunity dynamics. *Proceedings of the Royal Society B: Biological Sciences*, 278(1721):3042–3049.

Caswell, H. (1976). Community structure: a neutral model analysis. *Ecological monographs*, 46(3):327–354.

Cohen, J. E. (1970). A markov contingency-table model for replicated lotka-volterra systems near equilibrium. *The American Naturalist*, 104(940):547–560.

Connor, E. F. and Simberloff, D. (1979). The assembly of species communities: chance or competition? *Ecology*, 60(6):1132–1140.

Diamond, J. M. (1975). Assembly of species communities. In *Ecology and evolution of communities*, pages 342–444. Harvard University Press.

Dubart, M., Pantel, J. H., Pointier, J.-p., Jarne, P., and David, P. (2019). Modeling competition, niche, and coexistence between an invasive and a native species in a two-species metapopulation. *Ecology*, 100(6):e02700.

Faust, K. and Raes, J. (2012). Microbial interactions: from networks to models. *Nature Reviews Microbiology*, 10(8):538–550.

Forbes, S. A. (1907). *On the local distribution of certain Illinois fishes: an essay in statistical ecology*, volume 7. Illinois State Laboratory of Natural History.

Friedman, J. and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput Biol*, 8(9):e1002687.

Gotelli, N. J. (1991). Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. *The American Naturalist*, 138(3):768–776.

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81(9):2606–2621.

Gotelli, N. J. and Ulrich, W. (2012). Statistical challenges in null model analysis. *Oikos*, 121(2):171–180.

Hamelin, F. M., Allen, L. J., Bokil, V. A., Gross, L. J., Hilker, F. M., Jeger, M. J., Manore, C. A., Power, A. G., Rúa, M. A., and Cunniffe, N. J. (2019). Coinfections by noninteracting pathogens are not independent and require new tests of interaction. *PLoS Biology*, 17(12):e3000551.

Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, pages 210–221.

Hanski, I. and Gyllenberg, M. (1993). Two general metapopulation models and the core-satellite species hypothesis. *The American Naturalist*, 142(1):17–41.

Hanski, I. and Gyllenberg, M. (1997). Uniting two general patterns in the distribution of species. *Science*, 275(5298):397–400.

Hastings, A. (1980). Disturbance, coexistence, history, and competition for space. *Theoretical Population Biology*, 18(3):363–373.

Hastings, A. (1987). Can competition be detected using species co-occurrence data? *Ecology*, 68(1):117–123.

Hastings, A. (1991). Structured models of metapopulation dynamics. *Biological Journal of the Linnean Society*, 42(1-2):57–71.

Hastings, A. (2003). Metapopulation persistence with age-dependent disturbance or succession. *Science*, 301(5639):1525–1526.

Jentsch, A. and White, P. (2019). A theory of pulse dynamics and disturbance in ecology. *Ecology*, 100(7):e02734.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., et al. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology letters*, 7(7):601–613.

Levin, S. A. and Paine, R. T. (1974). Disturbance, patch formation, and community structure. *Proceedings of the National Academy of Sciences*, 71(7):2744–2747.

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*, 15(3):237–240.

MacArthur, R. H. and Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, pages 373–387.

Molina, C. and Stone, L. (2020). Difficulties in benchmarking ecological null models: an assessment of current methods. *Ecology*, 101(3):e02945.

Münkemüller, T., Gallien, L., Pollock, L. J., Barros, C., Carboni, M., Chalmandrier, L., Mazel, F., Mokany, K., Roquet, C., Smyčka, J., et al. (2020). Dos and don’ts when inferring assembly rules from diversity patterns. *Global Ecology and Biogeography*, 29(7):1212–1229.

Olivieri, I., Michalakis, Y., and Gouyon, P.-H. (1995). Metapopulation genetics and the evolution of dispersal. *The American Naturalist*, 146(2):202–228.

Opedal, Ø. H., von Numers, M., Tikhonov, G., and Ovaskainen, O. (2020). Refining predictions of metacommunity dynamics by modeling species non-independence. *Ecology*, 101(8):e03067.

Ovaskainen, O. and Abrego, N. (2020). *Joint species distribution modelling: with applications in R*. Cambridge University Press.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., and Abrego, N. (2017). How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology letters*, 20(5):561–576.

Pickett, S. T. and White, P. S. (2013). *The ecology of natural disturbance and patch dynamics*. Elsevier.

Slatkin, M. (1974). Competition and regional coexistence. *Ecology*, 55(1):128–134.

Sousa, W. P. (1984). The role of disturbance in natural communities. *Annual review of ecology and systematics*, 15(1):353–391.

Stone, L. and Roberts, A. (1990). The checkerboard score and species distributions. *Oecologia*, 85(1):74–79.

Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S., and Chase, J. M. (2020). A process-based metacommunity framework linking local and regional scale community ecology. *Ecology letters*, 23(9):1314–1329.

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M., Oksanen, J., and Ovaskainen, O. (2020). Joint species distribution modelling with the R-package Hmsc. *Methods in ecology and evolution*, 11(3):442–447.

Tilman, D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology*, 75(1):2–16.

Tilman, D., May, R. M., Lehman, C. L., and Nowak, M. A. (1994). Habitat destruction and the extinction debt. *Nature*, 371(6492):65–66.

Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22(2):252–260.

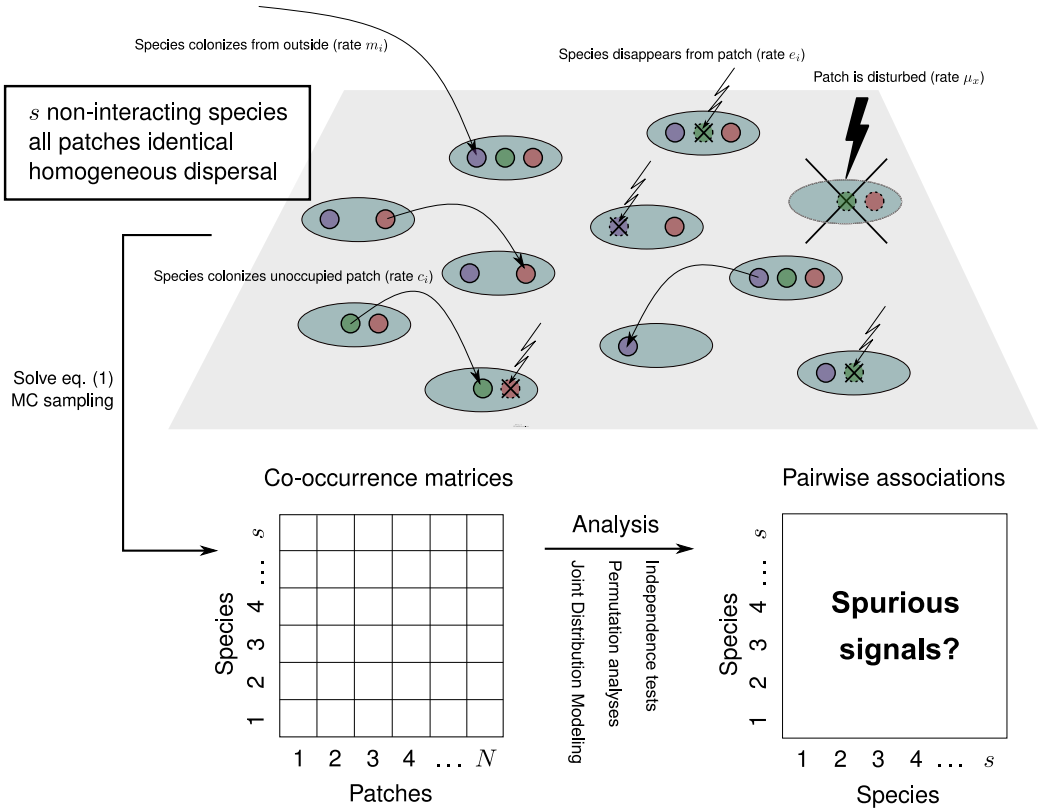

Figure 1: Graphical summary of our model and approach. Top: null metacommunity model, its assumptions and the processes it describes. Note that patch disturbances may occur at a certain rate, possibly dependent on the time since the last disturbance. Alternatively, they may occur after some prescribed amount of time (introducing a maximum life span of patches). Bottom: from the metacommunity model, using specific parameter values, we can generate sample co-occurrence matrices of arbitrary size. We can then subject these matrices to standard analytical methods to search for signals of species association. Considering that our model assumes no species interaction, no habitat heterogeneity or differences in habitat selection among species, we expect no significant association for any pair of species. Any such association would be a spurious signal (Type I error, or false positive).

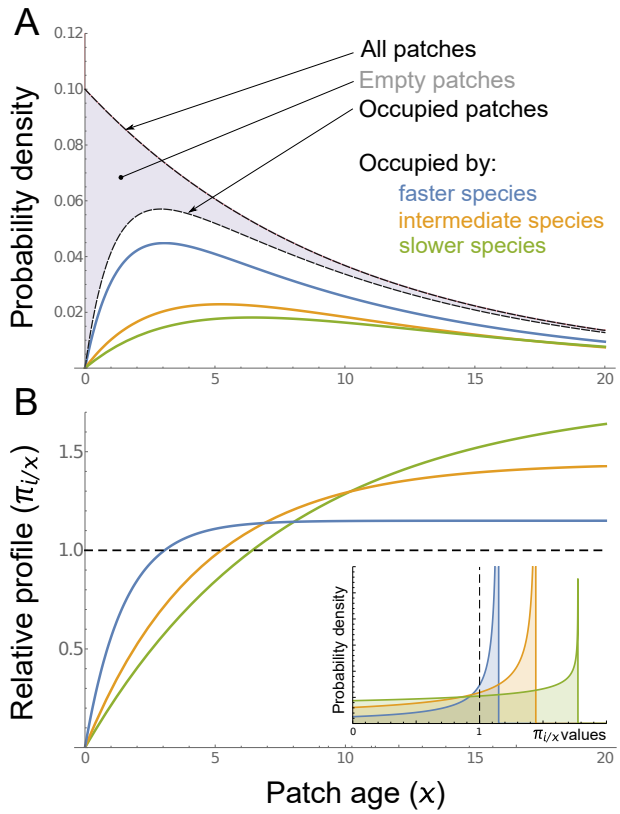

Figure 2: (a) Example equilibrium patch occupancies with three species differing in fastness. In this example the patch disturbance rate was $\mu = 0.1$ per unit time with no maximum age, so that the stable patch-age distribution was exponential with mean 10 (top curve). At each patch age, all three species have some occupancy $p_{i,x}$ (bottom three curves), leaving a certain fraction of patches empty. In this example the fastest species are also more abundant overall, but this needs not be the case. (b) The relative distribution profile of faster species has steeper initial slope and attains lower asymptotic values. As a species gets faster and faster, its profile converges to the horizontal dashed line, with value 1 for all patch ages. As shown in the insert, faster species have a smaller variance in $\pi_{i/x}$ values, and as a species gets faster and faster, the distribution converges to a Dirac delta function at value 1 (vertical dashed line). Parameters: $m_i = 0.01$, and $e_i = (0.05, 0.1, 0.2)$ and $c_i = (0.2, 0.3, 1)$ for slow, intermediate and fast species, resp.

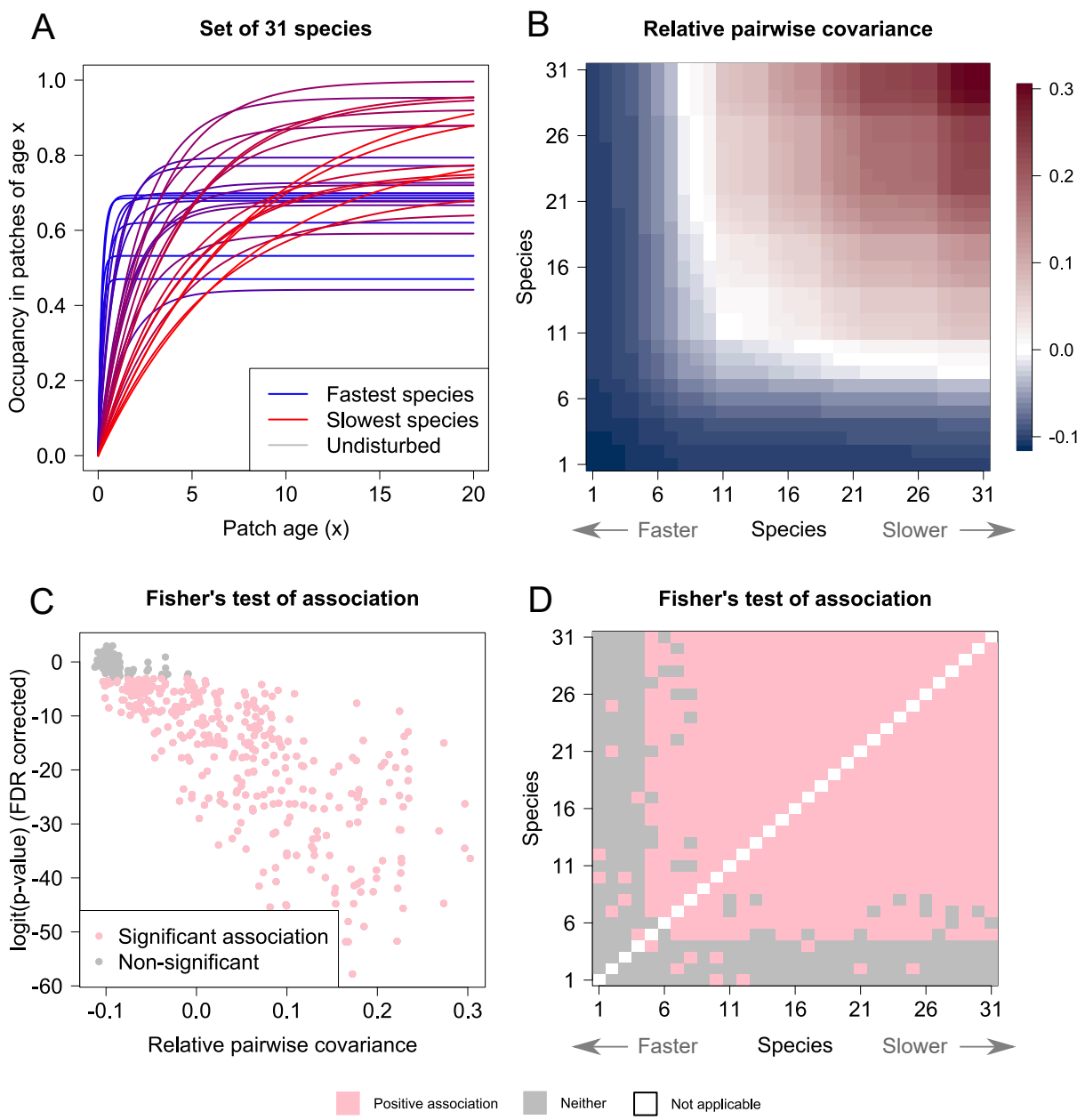

Figure 3: Testing for pairwise species independence. (A) Occupancy as a function of patch age x ($p_{i|x}$) for our synthetic community of 31 species. Species traits were drawn randomly in order to have variable overall occupancies and fastnesses. (B) Pairwise covariances of relative profiles for all species pairs. Species are ranked according to fastness: species 1 is the fastest (lowest variance), and species 31 is the slowest (highest variance). The variance of the average profile in the community ($\hat{\pi}_x$) was substracted from covariances to help distinguish lower/higher than average values (color bar). See equations (4) and (7). (C) Significance of Fisher's independence tests (logit p-values, FDR corrected), for every species pair, as a function of pairwise covariance. (D) Species pairs significantly associated according to Fisher's tests. Species are ordered as in (B). The sample co-occurrence matrix was generated from Eq. 1 with $N = 1000$ sites (patches) and the illustrative parameter set. See Supporting Information B for all details.

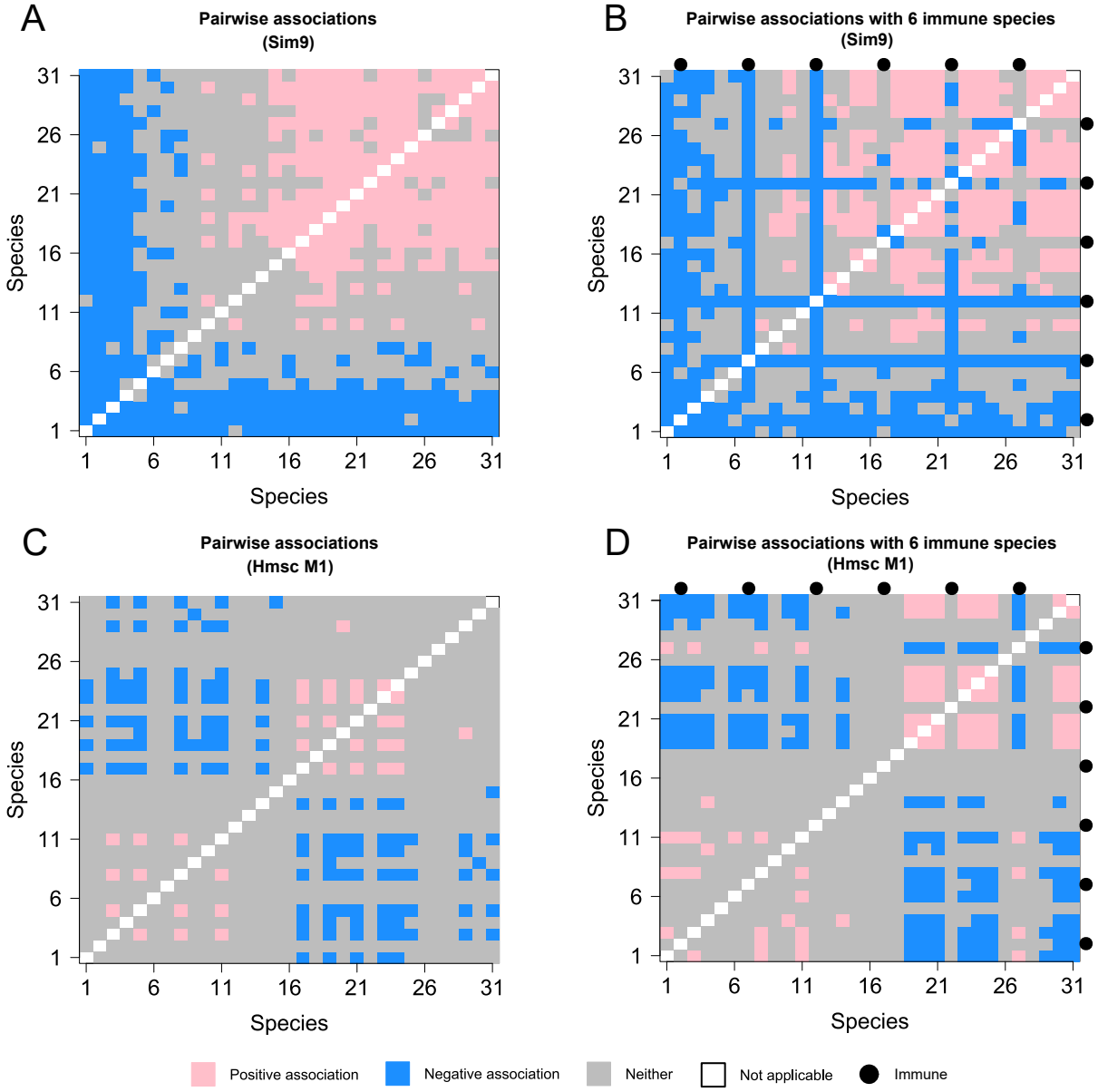

Figure 4: Controlling for patch richness. (A) Species pairs that are significantly associated according to the fixed-fixed (Sim9) permutation algorithm. (B) Same as (A), with a subset of species immune to patch disturbance (shown as bullets). (C) Species pairs significantly associated according to a joint species distribution model (Hmsc M1) that models patch richness as a single latent factor. (D) Same as (C), with a subset of species immune to patch disturbance. Same species and sample co-occurrence matrix as in previous figure.

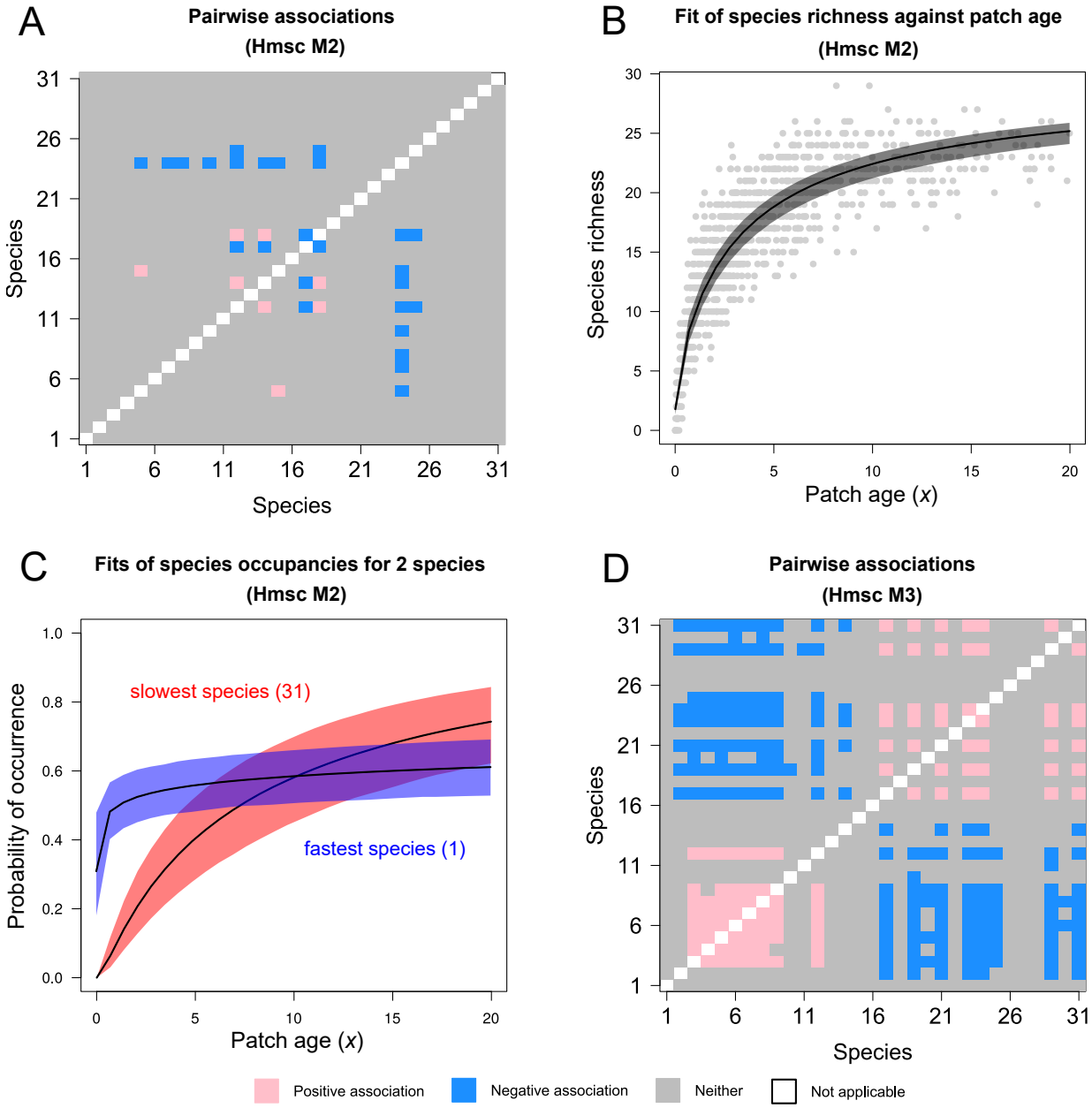

Figure 5: Using patch age as a covariate. (A) A joint species distribution model (Hmsc M2) that incorporates patch-age as an explicit environmental covariate yields correct results, even with patch disturbance. The few spurious species associations detected are within the type-I error rate. Note that log(patch age) was used as the covariate, since patch age is distributed exponentially. (B) Predicted species richness as a function of patch-age. Gray dots represent the actual data in the sample co-occurrence matrix. (C) Predicted probability of occurrence as a function of patch age, for the fastest species and the slowest species. (D) Same as (A) but taking species richness as a proxy for patch age: the joint species distribution model (Hmsc M3) fails in a similar way as Hmsc M1 did (Figure 4C). Same species and sample co-occurrence matrix as in previous figures.

Table 1: Summary of main findings. We only mention methods tested in this study.

Type of approach	**Example methods**	**Works?**	**Type of bias**
Direct independence tests	Fisher's tests; Sim2; Hmsc M0	No†	positive
Controlling for patch richness	Sim9; Hmsc M1; Hmsc M3	No*	positive/negative
Modelling effect of patch age	Hmsc M2	Yes	none

† only if no patch disturbance
* only if no patch disturbance, or all species similar (same fastness)

Supporting Information

- A. Mathematical supplements.
- B. Stochastic simulations.
- C. Supplementary Figures.