Journal of Animal Ecology

The relative importance of local and regional processes to metapopulation dynamics

Tad A Dallas[a,b,*], Marjo Saastamoinen[a,c], Torsti Schulz[a] and Otso Ovaskainen[a,d]

[a] *Organismal and Evolutionary Biology Research Programme, P.O. Box 65, 00014 University of Helsinki, Finland*

[b] *Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA*

[c] *Helsinki Institute for Life Sciences, University of Helsinki, Finland*

[d] *Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology. N-7491 Trondheim, Norway*

*Corresponding author: tad.a.dallas@gmail.com

Running title: Local & regional metapopulation drivers

Author contributions: TAD designed the study and performed the analyses. All authors contributed to manuscript writing.

Acknowledgments: We thank the coordinators and volunteers who participated in the Åland island survey since 1993. The research was funded by the Academy of Finland (grant 309581 to OO), the Research Council of Norway (SFF-III grant 223257), and the European Research Council (Independent Starting grant no. 637412 'META-STRESS' to MS). The Research Centre for Ecological Change is funded by the Jane and Aatos Erkko Foundation.

Conflict of interest: The authors have no conflicts of interest to declare.

Keywords: Patch occupancy, Spatial network structure, Species distribution, Metapopulation, Graph theory, Connectivity

Data accessibility: *R* code is available on figshare at `https://doi.org/10.6084/m9.figshare.7667096`.

Abstract

1. Metapopulation dynamics – patch occupancy, colonization, and extinction – are the result of complex processes at both local (e.g., environmental conditions) and regional (e.g., spatial arrangement of habitat patches) scales. A large body of work has focused on habitat patch area and connectivity (area-isolation paradigm). However, these approaches often do not incorporate local environmental conditions, or fully address how the spatial arrangement of habitat patches (and resulting connectivity) can influence metapopulation dynamics.

2. Here, we utilize long-term data on a classic metapopulation system – the Glanville fritillary butterfly occupying a set of dry meadows and pastures in the Åland islands – to investigate the relative roles of local environmental conditions, geographic space, and connectivity in capturing patch occupancy, colonization, and extinction. We defined connectivity using traditional measures as well as graph theoretic measures of centrality. Using boosted regression tree models, we find roughly comparable model performance among models trained on environmental conditions, geographic space, or patch centrality.

3. In models containing all of the covariates, we find strong and consistent evidence for the roles of resource abundance, longitude, and centrality (i.e., connectivity) in predicting habitat patch occupancy and colonization, while patch centrality (connectivity) was relatively unimportant for predicting extinction. Relative variable importance did not change when geographic coordinates were not considered and models underwent spatially-stratified cross validation.

4. Together, this suggests that the combination of regional scale connectivity measures and local-scale environmental conditions are important for predicting metapopulation dynamics, and that a stronger integration of ideas from network theory may provide insight into metapopulation processes.

Introduction

Species often occupy only a portion of potential habitat within their geographic range (MacArthur, 1984). This is especially true when species occupy small and fragmented habitats within a landscape, resulting in temporally dynamic occurrence across the set of interconnected habitat patches i.e., a metapopulation (Hanski, 1994a, 1999b). A large body of theory has emerged from the metapopulation concept at scales from examinations of entire metapopulations (Gotelli, 1991; Gilarranz & Bascompte, 2012), semi-independent networks (Hanski *et al.*, 2017), individual habitat patches (Ovaskainen & Hanski, 2003; Ovaskainen, 2017), and individuals within habitat patches (Ovaskainen & Hanski, 2004). The continued interest in metapopulations has produced many testable hypotheses concerning patch occupancy and dynamics (Ovaskainen & Saastamoinen, 2018), and how these quantities relate to metapopulation structure (Thomas, 1994; Hanski, 2001).

Naturally, there are many variables that interact to produce species occurrence in a given habitat patch (Guisan & Thuiller, 2005; Elith & Graham, 2009). However, despite this complexity, habitat patch area has emerged as a consistently good predictor of metapopulation dynamics (Hanski, 1994a; Thomas & Harrison, 1992; Hill, Thomas & Lewis, 1996). Habitat patch area, and associated area-isolation paradigm (Hanski, 1994a), has been linked to enhanced species persistence (Etienne, 2004) and colonization (Fleishman *et al.*, 2002), while also decreasing the probability of local extinction (Fleishman *et al.*, 2002; Hanski, 1994b; Day & Possingham, 1995). Much of this rests on the assumption that larger habitats can support larger populations and represent a larger target for incoming propagules from nearby patches (Ovaskainen & Saastamoinen, 2018); an assumption with mixed support (Bowman, Cappuccino & Fahrig, 2002; Rabasa, Gutiérrez & Escudero, 2008; Anderson & Meikle, 2010). Despite a focus on patch area, other variables are certainly related to metapopulation dynamics (Mortelliti, Amori & Boitani, 2010). For instance, the spatial position of habitat patches has been linked with patch occupancy (Ims, Petter Leinaas & Coulson, 2004; MacKenzie *et al.*, 2017), as patches in certain areas may be more likely to be colonized (or rescued) by immigration (Eriksson *et al.*, 2014). Spatial position may additionally serve as a proxy for some unmeasured aspect of habitat quality or environmental constraints on species occurrence. Further, local dynamics may be driven by ecological interactions and resource limitation, such that patch occupancy in a given habitat patch could be a result of interactions with competitors (Connor & Simberloff, 1979; Hamel *et al.*, 2013), resource limitation (Dennis & Eales, 1999; Dennis, Shreeve & Van Dyck, 2003), or natural enemies (Steffan-Dewenter & Schiele, 2008). Lastly, metapopulation dynamics could be a result of habitat patch connectivity driven by physical distance of the patches and/or by the dispersal ability of the focal species, suggesting that spatial network statistics may explain patch occupancy (Gilarranz & Bascompte, 2012; Grilli, Barabás & Allesina, 2015).

Measures of habitat patch importance in spatial networks have been developed largely outside of the realm of metapopulation ecology, despite measuring similar – and sometimes equivalent – properties (see (Urban *et al.*, 2009) and Box 1). So what benefit do we obtain from using measures from graph theory in place of, or in addition to, existing measures of the importance of a habitat patch to the metapopulation, such as patch contribution to metapopulation capacity (Ovaskainen & Hanski, 2003)? Measures of centrality attempt to quantify flow of information or individuals between habitat patches, but centrality itself can be measured in many different ways. That is, measures can be quite local (focused only on the immediate connections of a given habitat patch with other patches in the immediate vicinity) or global (incorporating information on the spatial distribution of all habitat patches in the network and the connections between them). This is advantageous as ecological processes may occur at both of these scales simultaneously. One clear example of the potential benefits of using graph theoretic centrality measures in place of existing connectivity measures is in the case of 'stepping stone' habitat patches (Bodin & Saura, 2010), which serve to connect two habitat patches which otherwise would not be connected by dispersal. In graph theory, betweenness centrality measures the number of shortest paths between all pairs of habitat patches in the network which go through a given habitat patch. This essentially measures, at the network scale, the importance of a habitat patch as a potential stepping stone. The further integration of metapopulation ecology and graph theory will greatly advance our understanding of metapopulation dynamics (Urban *et al.*, 2009).

But how important are measures of connectivity – either from graph theory or from metapopulation ecology – relative to aspects of habitat patch quality, spatial position, or patch area? Numerous studies have explored the relationship of each of these factors to metapopulation dynamics (e.g., Hanski (1994a); Fleishman *et al.* (2002); Prugh *et al.* (2008)), but few have weighed the relative effects of different covariate groups (but see Rabasa, Gutiérrez & Escudero (2008); Fleishman *et al.* (2002)). Understanding the relative importance of each of these variable sets on metapopulation dynamics is a pressing need, as some things change (local environmental conditions) and some things tend to stay the same (spatial arrangement of habitat patches). Failing to account for this could lead to inaccurate predictions concerning metapopulation persistence or misidentification of habitat patch conservation targets. It is also important to note that habitat patch quality, spatial position, and habitat patch centrality – which putatively determine metapopulation dynamics – likely interact to produce spatial variation in habitat patch occupancy, colonization, and extinction processes. For instance, resource limitation may only control patch occupancy in a given habitat if enough individuals are present and able to disperse to the habitat. This density-dependence would result in an interaction between resource availability and habitat patch isolation. As such, approaches capable of estimating the relative importance of local (e.g., environmental conditions) and regional (e.g., spatial arrangement of patches) factors are needed to advance our understanding of metapopulation dynamics.

Here, we address two current shortcomings in examinations of metapopulation dynamics. First, we provide a clear link between graph-theoretic measures (i.e., centrality) to connectivity as defined in metapopulation ecology. Second, we examine the relative influence of geographic position, habitat (e.g., resource availability), and patch connectivity on metapopulation dynamics. To do this, we utilize data from a classic well-studied ecological metapopulation, the Glanville fritillary metapopulation in the Åland islands (Ojanen *et al.*, 2013; Hanski *et al.*, 2017). While numerous studies have examined the influence of patch-level *or* network-level covariates on metapopulation processes, weighing the relative importance of different covariate groups is far more rare, despite the potential for synergistic effects (see Table 1). Our aim is to quantify the contributions of patch area, spatial location, local habitat-level variables, and connectivity (i.e, patch centrality in the dispersal network) on patch occupancy (fraction of times a patch was occupied), colonization, and extinction. In doing so, we highlight the similarities between measures of connectivity and centrality (Box 1), and explore whether measures derived from metapopulation theory and graph theory are correlated, or whether they measure fundamentally different aspects of the network properties (Minor & Urban, 2007; Urban *et al.*, 2009). Further, we provide evidence for the importance of local habitat conditions and connectivity in driving metapopulation dynamics, suggesting that the combination of local environmental conditions with measures of dispersal connectivity may best explain metapopulation dynamics (see Table 1). The continued integration of graph theoretic measures and flexible statistical approaches that allow estimation of relative variable importance will enhance our understanding of the relative roles of geography, environment, and dispersal to metapopulation dynamics.

Box 1: Linking metapopulation statistics and graph theory

The development of theory related to metapopulations and spatial graphs – despite the striking similarities in application – has been largely separate (but see (Urban *et al.*, 2009; Dale & Fortin, 2010)). This has lead to the development of statistics different in name, but identical (or quite similar) in application. For instance, habitat patch connectivity (S_i; Equation 1) is a measure from metapopulation ecology, and quantifies the total immigration potential into a given habitat patch (Hanski, 1999a). This considers the receiving patch area scaled by some constant im, a negative exponential dispersal kernel ($e^{-\alpha d_{ij}}$), and the influence of the donor patch area raised to an emigration term (A_j^{em}).

$$S_i = \sum_{j \neq i} A_i^{im} e^{-\alpha d_{ij}} A_j^{em} \tag{1}$$

If we consider the links between habitat patches in the spatial network as potential dispersal pathways, the edge between two patches in the network can be defined according to that same negative exponential dispersal kernel, and patch area can be included in these link weights if the influence of habitat patch area on immigration and emigration is well understood (Hanski *et al.*, 2017). Then, a measure from graph theory, weighted degree centrality (sometimes referred to as *strength*) is quantified by summing the edges going into a given habitat patch. This is equivalent to connectivity measures as developed in metapopulation ecology, dependent on how patch area is incorporated, and whether degree centrality is calculated on a directed graph (i.e., dispersal pathways between two nodes are non-equal). Further, degree centrality is not the only form of centrality in graph theory, and each different formulation of centrality captures some unique aspect of centrality. Degree centrality inherently captures local dynamics, as it is concerned with direct connections of a given habitat patch. However, other measures utilize information on the entire network and connections between other nodes. For example, betweenness centrality measures the importance of habitat patches as bridges between other habitat patches, which is important to conservationists and managers when designing reserves, especially for migratory species (Fall *et al.*, 2007). Further, betweenness centrality may better capture the tendency for patches to maintain connections between patches too far apart to be connected. Meanwhile, closeness centrality, which measures the mean shortest path distance between a patch to all other habitat patches, may captures spatial aggregation of habitat patches, with the potential to be a better predictor of metapopulation dynamics than more local measures of connectivity (e.g., degree centrality).

Another example of this is the close relationship between the contribution of a habitat patch to overall metapopulation capacity (λ_i), developed in the study of metapopulations, and eigenvector centrality from graph theory. While not directly analagous, both use a eigenvector decomposition of the dispersal network to estimate the importance of each habitat patch to the overall structure of the spatial network. Using the Åland metapopulation as an example, we see the clear positive relationship between habitat patch contribution to metapopulation capacity

(Ovaskainen, 2003; Grilli, Barabás & Allesina, 2015) and eigenvector centrality (Figure 1). A more direct example, though less often used currently, is the hub score (Kleinberg, 1999), which is nearly identical to metapopulation capacity. The only difference is that metapopulation capacity is calculated on the dispersal matrix ($\mathbf{M}$) and the hub score is calculated on the positive definite matrix obtained by multiplying the matrix by it's transpose ($\mathbf{M} \times \mathbf{M^T}$).

The theory developed for the study of networks – even solely the development of theory related to spatial graphs – is more general and more broadly utilized than the theory of metapopulations (Newman, 2003; Barthélemy, 2011), despite the fact that metapopulations are clear examples of spatial graphs. The application of approaches from graph theory may provide further insight into metapopulation structure and resulting metapopulation dynamics.

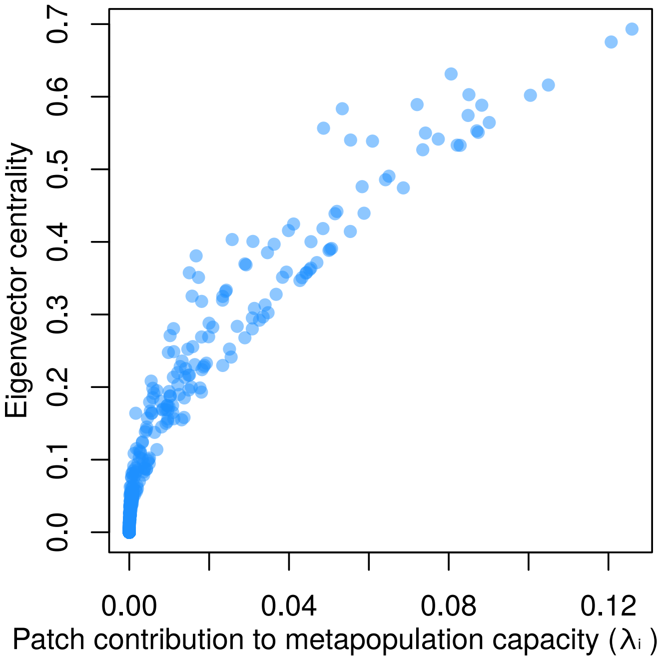

Figure 1: A strong positive relationship exists between patch contribution to metapopulation capacity (λ_i) and eigenvector centrality. Each point corresponds to a habitat patch in the Åland island metapopulation system. Eigenvector centrality was based on a dispersal network formed assuming an exponential decay in dispersal probability between patches ($\alpha = 1$ and p= 0.001, as described further in the *Spatial network formation* section).

Methods

Glanville fritillary metapopulation

In the Åland islands, a set of nearly 5000 habitat patches have been monitored annually since 1993. The habitat in the Åland Islands is highly fragmented and the butterfly has a classic metapopulation structure with a high rate of population turnover – i.e., extinctions and re-colonizations (Hanski, 1999b, 2011). However, as some habitat patches were not surveyed for the entire duration of the study, we restrict our analyses to patches surveyed between 2000 - 2017, resulting in a total of 4652 habitat patches distributed broadly across the Åland islands. However, for patches for which environmental data was available prior to 2000, we included these years to estimate the mean environmental conditions. Each habitat patch is a dry meadow or pasture occupied by one or more host plant species – *Plantago lanceolata* or *Veronica spicata* – which serve as a larval food source and oviposition resource to the butterfly of interest, *Melitaea cinxia.*

The dry meadows and pastures have been surveyed for the presence and numbers of larval groups during fall (Hanski, 1999b, 2011). This is possible as the females of the Glanville fritillary butterfly lay clutches of eggs, the larvae live gregariously, and at the end of the summer the larvae build a conspicuous "winter nest" at the base of the host plant inside which they diapause overwinter in groups of mainly full sibs (Kuussaari *et al.*, 2004; Fountain *et al.*, 2018). Each fall all of the potential habitat patches are surveyed for the presence of these larval nests (see (Ojanen *et al.*, 2013) for details of the survey). Based on control surveys it has been estimated that the presence of the butterfly is not detected in up to 15% of occupied patches with non-detection mainly occurring in very small populations (Hanski *et al.*, 2017). Based on the long term data we know that all local populations are more or less ephemeral, due to being very small and commonly having just a single or a few larval groups in a given year (Hanski, 1999b, 2011).

Patch occupancy, colonization, and extinction

Occupancy was quantified as the fraction of times a habitat patch was occupied by *M. cinxia* during the survey. This provides insight into how often a given habitat patch contributed to metapopulation dynamics, as more frequently occupied patches are likely more important to enhancing metapopulation persistence and providing propagules to other nearby patches. The spatial distribution of patch occupancy clearly identifies hotspots of habitat patches which maintain the metapopulation (Figure 2).

Colonization rate captures how fast a habitat patch becomes re-colonized after a local extinction. We quantified colonization probability as the number of times that *M. cinxia* was present when it did not occur in the previous sampling period divided by the total number of possible colonization events (i.e., the number of sampling periods where the species was absent, not considering the most recent sampling period). Extinction probability was measured in a similar manner, calculated as the number of times a species was recorded as absent when it was observed in that patch in the prior sampling period, divided by the total number of potential extinction events. Patches with high turnover – those that are colonized and go extinct often – may simply be sinks for propagules from more persistent patches. On the other hand, these patches may contribute strongly to metapopulation persistence if they serve as temporary spillover habitats or provide dispersal connections with more distant patches (Howe, Davis & Mosca, 1991; Hanski & Simberloff, 1997).

The full number of habitat patches ($n = 4652$) was used for analyses of patch occupancy. Habitat patches that were never occupied ($n = 2595$), and those that remained occupied for the entire sampling duration ($n = 21$) were removed from calculation of colonization and extinction, resulting in 2057 and 4631 habitat patches, for examinations of colonization and extinction, respectively.

Defining the spatial network

Habitat patches exist in a mosaic of inhospitable habitat to *M. cinxia*, and links between habitat patches represent potential dispersal pathways. Based on previous research (Hanski *et al.*, 2017) we considered dispersal probability to decay exponentially with geographic distance between habitat patches. We constructed a network based on this exponential decay ($\alpha = 1$ km^{-1}), and removed links below a threshold dispersal probability ($p = 0.001$). We examine the sensitivity of the resulting dispersal network structure in the Supplemental Material, finding no appreciable difference in patch connectivity estimates (see Figure S1). Patch area may influence dispersal probability and subsequent links between habitat patches in the network (Hanski, 2001; Hanski *et al.*, 2017). We incorporated the influence of patch area on the structure of the dispersal network by modifying the negative exponential dispersal kernel, where links between two habitat patches were defined as a function of the area of both patches (A_i and A_j), both of which were raised to constants obtained from previous studies (Hanski *et al.*, 2017), which represent the relationships between patch area and immigration ($im = 0.3$) and emigration ($em = 0.3$) rates (see Equation 1). This is discussed further in Box 1, which conceptually links measures of centrality to existing concepts in metapopulation ecology. We found qualitatively similar results when habitat patch area was not allowed to influence dispersal links (see Supplemental Materials).

Variables influencing patch occupancy, colonization, and extinction

We divided variables into four different groups, in order to compare model performance among variable groups, while also considering a full model including all variables. We also consider every combination of the variable groups in the Supplemental Materials, providing even further support for our conclusions. The variable groups consisted of *patch area* (a baseline model which only considers the log-transformed habitat patch area), *habitat* (containing local patch level environmental variables), *spatial* (containing spatial position of each habitat patch), and *network* (containing measures of patch centrality). Expected relationships between variable groups and metapopulation dynamics are provided in Table 1, and each of the variable groups are outlined in Table 2, with each variable described below.

Patch area was estimated during sampling, with the median patch area being approximately 0.6 ha. The spatial location of each habitat patch was mapped with GPS during the survey (Ojanen *et al.*, 2013). Grazing pressure was estimated as the estimated fraction of the habitat patch subjected to grazing pressure based on observations of damaged plants or the presence of grazers (e.g., ungulates). We quantified resource availability as the mean abundance, and the summed mean abundance of the two host plants (*Plantago lanceolata* and *Veronica spicata*), where abundance of each host plant was estimated based on an ordinal scale between 0 and 3, with larger values corresponding to a greater plant abundance. Previous findings in a rodent herbivore metapopulation suggest that temporal variability in resources can influence metapopulation dynamics (Fernández, Román & Delibes, 2016). We explore this in the Supplemental Materials by calculating the standard deviation in total resource availability (the summed abundance of both host plants). We find little evidence that variability in resource abundance influences metapopulation dynamics (see Supplemental Materials for further analyses and discussion), suggesting that species life history may play a large role in estimating the relative importance of spatial and environmental variables on metapopulation dynamics (Fernández, Román & Delibes, 2016). Resource quality may be reduced as a function of infection by a powdery mildew pathogen, which has been found to reduce *M. cinxia* larval development over the summer (Rosa *et al.*, 2018) and influence overwintering survival (Laine, 2004). Mildew infection was estimated by quantifying the fraction of times mildew pathogen was detected in each habitat patch.

Habitat patch importance in the spatial network was estimated using patch centrality measures. Specifically, we examined four common centrality measures, each capturing different aspects of habitat patch importance in the dispersal network (**M**; equations for each connectivity measure are provided in the Supplemental Materials). First, weighted degree centrality – also called *strength* – measures the summed links (dispersal pathways) for each habitat patch. This measures the immediate connections to neighboring patches. Next, we considered closeness centrality, which incorporates the structure of the overall network, measuring the average shortest path distance between each habitat patch to all other habitat patches. Habitat patches with large closeness values would be well connected to other patches in the context of the entire network, while degree centrality measures habitat patch importance in a neighborhood context. Next, we considered betweenness centrality, which measures the number of shortest paths between habitat patches that go through a given habitat patch. This is important, as habitat patches with high betweenness may serve as stepping stones between two otherwise unconnected habitats. Lastly, we measured eigenvector centrality, which measures the importance of habitat patches as defined by the importance of connected habitat patches. That is, a habitat patch may not be strongly connected to many other habitats, but be connected to a patch that is quite well connected to other patches (i.e., serves an important role in the metapopulation). This could occur when a patch is spatially removed from much of the spatial network, but connected to nearby patches which are more well connected to other habitat patches.

Boosted regression tree models

Boosted regression tree (BRT) models were used to assess how patch area, geographic space, habitat-level variation, and patch centrality influence *M. cinxia* occupancy and colonization using the `gbm` R package (with contributions from others, 2017). This modeling approach has been used previously for prediction (Elith, Leathwick & Hastie, 2008; De'Ath, 2007), in part because it allows for non-linear responses and variable interactions. Since the regression tree is hierarchical, "upstream" splits based on one variable influence "downstream" splits, which automatically models variable interactions. Further, the process of *boosting* enhances learning on complex data, as the process produces many regression trees with a small number of splits, each of these "weak learners" iteratively build on previous trees to account for the remaining variation. This approach removes the need to partition variance among submodels, as the goal is not to examine the components of variance explained, but to assess overall model performance with the inclusion or exclusion of particular variable sets.

For each of the four covariate groups and the full model containing all covariates, models were trained, cross-validated, and evaluated for performance five times (each on a different random subset of 80% of the data) to examine the consistency of model performance and covariate relative importance. Models were trained using a maximum of 50,000 trees, with a learning rate of 0.001 (Elith, Leathwick & Hastie, 2008), Gaussian error structure, and an interaction depth of 3, which allows for interactions between covariates. All models were internally cross-validated (5-fold) to determine the optimal number of regression trees.

Models were trained on 80% of the data, and the remaining 20% was used to assess model performance. Accuracy was quantified using Spearman's rank correlations between predicted values from the trained model and the empirical estimates of occupancy, colonization or extinction for each habitat patch in the 20% of the data which was used for testing (i.e., those data that were not used for model training). In the Supplemental Materials, we further quantify accuracy using Pearson's correlation and root mean square error (RMSE).

It is possible that spatial autocorrelation in metapopulation dynamics could lead to model overfitting when trained on spatial coordinate data. This would inflate the relative contribution of latitude and longitude in the full models, and lead to the spatial submodel appearing to perform well, when in fact it is simply fitting to spatial variation. While this could be informative if system-specific prediction was the goal, the ability of the model to extrapolate would be compromised. To explore the effect of spatial predictors on model transferrability, we also performed the cross validation by dividing the data spatially into five longitudinal folds (models were trained on four, and used to predict the remaining data).

The relative importance of each predictor variable in the full model containing all the covariates was estimated by quantifying the relative improvement to model fit as a result of the inclusion of a given covariate into the model, weighted by the number of trees in which the covariate occurred (Elith, Leathwick & Hastie, 2008; De'Ath, 2007). The resulting relative contribution values are scaled between 0 and 100, with larger numbers corresponding to higher variable importance, and the relative importance of all covariates summing to 100. To assess how important covariates influenced model predictions, we examined partial dependence plots, which capture the influence of a given variable on occupancy or colonization after accounting for other covariates (Elith, Leathwick & Hastie, 2008). Data and code to reproduce the analyses is provided at `https://doi.org/10.6084/m9.figshare.7667096`.

Results

Boosted regression tree model performance differed as a function of covariate group, with models trained on patch area generally performing the worst, and the model including all covariates performing best (Figure 3). The remaining models – consisting of local habitat variables, geographic location, or patch centrality – performed approximately equivalently (Figure 3). Considering all combinations of submodels, we find that the full model typically performed best, though in some cases the inclusion of patch area in the full model actually reduces model performance slightly, as does the geographic coordinates of the habitat patches (see Supplemental Materials). This suggests that the most important covariate sets to estimating metapopulation dynamics are local environmental conditions and habitat patch centrality (connectivity) measures (see Supplemental Materials for an expanded discussion). Model performance generally decreased when data were spatially stratified during 5-fold cross validation (open circles in Figures 3 and 4), suggesting the existence of a spatial signal in patch area, habitat characteristics, and spatial network structure. This spatial signal could exist through spatial autocorrelation, or because the effect of the covariate on metapopulation dynamics differs across space. Despite the existence of a spatial signal that influenced all submodels (e.g., the *habitat* model in Figure 3), the model including all variables tended to still outperform the submodels, and relative variable importance in these models was essentially unchanged by the cross-validation approach (Figure

4). However, the *habitat* model tended to perform just as well as the full model when models were spatially cross-validated, suggesting the importance of the local habitat on metapopulation dynamics (Figure 3).

Model performance and ranking were insensitive to the measure of model performance used (see Supplemental Materials). For models of extinction probability, the model containing local habitat covariates performed quite well, and submodels were relatively unaffected by the spatially stratified cross validation (Figure 3). Together, our findings suggest that patch occupancy, colonization, and extinction may be estimated to an approximately equal extent from detailed data on local habitat patch quality (*habitat* model) or more regional measures of patch connectivity (*network* model), but that joint effects between variables necessitate the inclusion of both local scale habitat variables and regional scale patch connectivity.

Variables influencing patch occupancy, colonization, and extinction

While many of the models trained on different covariate groups performed nearly equivalently (Figure 3), the relative importance of covariates in the full model under random cross-validation suggests that resource availability, longitude, and degree centrality were the dominant contributors to model performance (Figure 4). When latitude and longitude were not included in the spatially cross-validated models, the key predictors remained quite similar (i.e., resources and degree centrality). Eigenvector centrality, a measure of connectivity which incorporates information on connections of patches which a focal node is connected to, became more important in the spatially cross-validated models, potentially as a result of this measure capturing aspects of the spatial positions of the habitat patches. However, eigenvector and degree centrality tend to be highly correlated ($r = 0.48$, $p < 0.001$), and are both similarly related to metapopulation dynamics (Figure S6).

The partial dependence plots of each covariate in the full model suggest that resource availability and degree centrality both were positively related to occupancy and colonization (Figure S6). However, while resource availability was important and non-linearly related to extinction probability prediction, no measure of patch centrality (connectivity) improved the model substantially. The importance of patch centrality to patch occupancy and colonization relates to the amount of immigration to a given patch, which is naturally related to patch colonization probability (Hanski, 1991, 1999b), and could also reduce extinction risk through rescue effects (Eriksson *et al.*, 2014; Ovaskainen, 2017). However, this effect appeared weak, as models of extinction containing patch centrality only marginally outperformed a model containing only patch area (Figure 4), and no patch centrality measure was in the top three predictive variables in the full model (Figure S6). When patch area was not allowed to influence patch centrality measures, patch area became more important in estimating metapopulation dynamics. However, patch centrality measures still retained an important role in estimating metapopulation dynamics as well (see Supplemental Materials).

Interestingly, the summed resource abundance was more important than the abundance of either host plant (*P. lanceolata* and *V. spicata*) in isolation, suggesting the importance of considering the entire resource community instead of simply the most dominant host plant (*P. lanceolata*). Further, this value of resource abundance was the top predictor in all three full models of patch occupancy, colonization, and extinction (Figure 4), suggesting a pronounced effect of resource availability on metapopulation dynamics. The stronger relative effect of total resource abundance instead of the abundance of either host plant may relate to variable feeding preferences of individuals in a population, or behavioral flexibility in host plant utilization. That is, even if both resource plants were equally suitable resources, low abundance of one resource does not negate the presence of another suitable resource, making the summed resource abundance a clearer measure of resource availability for the butterflies.

Discussion

Metapopulation dynamics were best captured when both local environmental conditions and regional scale effects of habitat patch arrangement were considered. Secondly, while degree centrality – which is equivalent to how connectivity is typically defined in metapopulation studies – was largely the most important connectivity measure, other connectivity measures which incorporate more information about the surrounding network were also important (e.g., eigenvector centrality in models of occupancy and colonization). Together, this suggests that future research should incorporate multiple scales of information to understand metapopulation dynamics. Further, the joint effects of local and regional variables served to enhance model prediction, as evidenced by the substantial improvement in the full model relative to models including habitat, spatial, or network variables separately. Models incorporating local habitat variables, patch centrality, and geographic location performed nearly equivalently in estimating metapopulation dynamics, suggesting that the performance of more ecologically meaningful (habitat variables) models was roughly equivalent to less ecologically meaningful (spatial patch location) models. This is potentially due to systematic spatial variation in patch quality, the existence of strong dispersal limitation, or simply a model overfit to spatial data (see Supplemental Materials). Weighing the relative importance of all covariates in the full model, we consistently found that resource availability and degree centrality were important in estimating patch occupancy, colonization, and extinction probability (though patch area was comparably as important as patch centrality for extinction probability estimation). While network statistics may provide equivalent performance as more system-specific covariates for predicting patch occupancy and colonization, it is the combination of spatial processes, resource availability (Hanski *et al.*, 2017), and patch centrality (connectivity) that, in concert, best capture overall metapopulation dynamics.

The relative importance of network statistics to model performance suggests that metapopulation dynamics are strongly influenced by the structure of the network of habitat patches and the dispersal connections between them. This supports previous findings that patch centrality, independent of habitat patch quality, can approximate patch occupancy patterns (Hanski, 1991, 2011). However, these studies have largely focused on the role of patch area as it influences centrality, a connection which may take a variety of functional forms (Anderson & Meikle, 2010; Hambäck & Englund, 2005) given density-dependence in dispersal processes. We find that excluding the influence of patch area on centrality measures does tend to increase the influence of patch area estimates relative to patch centrality, and reduces the predictive accuracy of the *network* submodel greatly, suggesting that taking patch size when estimating dispersal connections between habitat patches is important (see Supplemental Materials). By the same token, the importance of resource availability suggests an important role for local patch quality on metapopulation dynamics, and the importance of habitat patch geographic position suggests that dispersal limitation and historical patch occupancy can influence resulting metapopulation dynamics. Lastly, the relative unimportance of patch connectivity to extinction probability may provide a further signal of the importance of scale, as occupancy and colonization may be more dependent on regional scale processes connecting habitat patches to one another, while extinction may be far more dependent on local environmental conditions, such as resource availability (Franzén & Nilsson,

2010) (but see (Rabasa, Gutiérrez & Escudero, 2008)). That is, while connectivity may rescue populations from extinction, patch extinction probability may ultimately be more a function of local environmental conditions than patch connectivity.

Apart from considering both local patch-scale processes and regional processes simultaneously, it is important to consider how dynamic or successional habitats can influence metapopulation dynamics (Hodgson, Moilanen & Thomas, 2009). That is, patch occupancy, colonization, and extinction were calculated under the assumption that the habitat did not change substantially, and that mean quantities accurately captured patch quality. We partially addressed the issue of dynamic environments by considering variation in resource abundance, which was found to be unimportant to estimating metapopulation dynamics (see Supplemental Material). Apart from dynamic habitats, numerous layers of complexity have been added to the existing patch area - connectivity paradigm, including incorporating informed or aggregated dispersal (Conradt *et al.*, 2000; Smith & Peacock, 1990), matrix habitat quality (Kuussaari, Nieminen & Hanski, 1996; Ricketts, 2001), and genetic information (Fountain *et al.*, 2018; Lamy *et al.*, 2012). The question then becomes, which of these additional layers are among the most important? If prediction of patch occupancy, colonization, and extinction is equally possible using data on spatial position compared to models incorporating patch level habitat variation or genetic data, it seems worthwhile to assess both the reasons behind the similarity, as well as the overall goal of the research. That is, additional layers become unnecessary if prediction of metapopulation dynamics is the goal, as simple measures of habitat patch centrality – even in the absence of habitat patch area – predict dynamics comparably to more highly parameterized models incorporating patch-level covariates. This is not to say that future research on the environmental, spatial, and genetic factors affecting metapopulation dynamics is not warranted. Quite the contrary. However, it would be useful to weigh the effect of these additional layers relative to basic models incorporating only information on patch area or network structure, as these simple models can provide benchmarks to assess the relative importance of additional factors.

Metapopulation ecology shares numerous conceptual and analytical commonalities with landscape ecology (DiLeo, Husby & Saastamoinen, 2018; Howell *et al.*, 2018) and network ecology (Box 1 and (Urban & Keitt, 2001; Urban *et al.*, 2009)). Bridging these disciplines can provide conceptual synthesis and lead to a better understanding of patch occupancy patterns (Urban & Keitt, 2001; Rozenfeld *et al.*, 2008; Gilarranz & Bascompte, 2012; Zamborain-Mason *et al.*, 2017). We find that local scale habitat variables are equally capable of predicting metapopulation dynamics as regional scale measures of connectivity, but that the best performing models included both local and regional scale variables together. This result may not scale to other metapopulation systems. However, differences in the balance of local scale patch quality and regional scale patch connectivity in other systems may provide insight into the drivers of metapopulation dynamics. Species life history becomes important to consider as well, as the dynamics of a species with narrow environmental tolerance and large dispersal kernel will be much more controlled by local scale processes than regional connectivity. Overall, our findings suggest that multi-scale approaches to estimating patch occupancy are important, especially considering the use of patch occupancy models in conservation decisions (Lande, 1988; Hanski & Ovaskainen, 2000; Lipcius *et al.*, 2008). Lastly, the use of statistical tools allowing for non-linear relationships and variable interactions is important to weighing the relative variable importance. A focus on the ability to predict metapopulation dynamics is paramount given shifting environmental conditions and land use changes resulting in non-random habitat patch destruction, deterioration, and alteration to dispersal links among habitat patches.

References

Alexander, H.M., Foster, B.L., Ballantyne, F., Collins, C.D., Antonovics, J. & Holt, R.D. (2012) Metapopulations and metacommunities: combining spatial and temporal perspectives in plant ecology. *Journal of Ecology*, **100**, 88–103.

Anderson, C.S. & Meikle, D.B. (2010) Genetic estimates of immigration and emigration rates in relation to population density and forest patch area in Peromyscus leucopus. *Conservation Genetics*, **11**, 1593–1605.

Austin, D. (2006) How Google finds your needle in the webs haystack. *American Mathematical Society Feature Column*, **10**, 12.

Barthélemy, M. (2011) Spatial networks. *Physics Reports*, **499**, 1–101.

Bodin, Ö. & Saura, S. (2010) Ranking individual habitat patches as connectivity providers: integrating network analysis and patch removal experiments. *Ecological Modelling*, **221**, 2393–2405.

Bowman, J., Cappuccino, N. & Fahrig, L. (2002) Patch size and population density: the effect of immigration behavior. *Conservation ecology*, **6**.

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

Conradt, L., Bodsworth, E., Roper, T. & Thomas, C. (2000) Non-random dispersal in the butterfly Maniola jurtina: implications for metapopulation models. *Proceedings of the Royal Society of London B: Biological Sciences*, **267**, 1505–1510.

Dale, M. & Fortin, M.J. (2010) From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*, **41**.

Day, J.R. & Possingham, H.P. (1995) A stochastic metapopulation model with variability in patch size and position. *Theoretical Population Biology*, **48**, 333–360.

De'Ath, G. (2007) Boosted trees for ecological modeling and prediction. *Ecology*, **88**, 243–251.

Dennis, R.L. & Eales, H.T. (1999) Probability of site occupancy in the large heath butterfly Coenonympha tullia determined from geographical and ecological data. *Biological Conservation*, **87**, 295–301.

Dennis, R.L., Shreeve, T.G. & Van Dyck, H. (2003) Towards a functional resource-based concept for habitat: a butterfly biology viewpoint. *Oikos*, pp. 417–426.

DiLeo, M., Husby, A. & Saastamoinen, M. (2018) Landscape permeability and individual variation in a dispersal-linked gene jointly determine genetic structure in the Glanville fritillary butterfly.

Elith, J. & Graham, C.H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Eriksson, A., Elías-Wolff, F., Mehlig, B. & Manica, A. (2014) The emergence of the rescue effect from explicit within-and between-patch dynamics in a metapopulation. *Proc. R. Soc. B*, **281**, 20133127.

Etienne, R.S. (2004) On optimal choices in increase of patch area and reduction of interpatch distance for metapopulation persistence. *Ecological Modelling*, **179**, 77–90.

Fall, A., Fortin, M.J., Manseau, M. & OBrien, D. (2007) Spatial graphs: principles and applications for habitat connectivity. *Ecosystems*, **10**, 448–461.

Fernández, N., Román, J. & Delibes, M. (2016) Variability in primary productivity determines metapopulation dynamics. *Proc. R. Soc. B*, **283**, 20152998.

Fleishman, E., Ray, C., Sjögren-Gulve, P., Boggs, C.L. & Murphy, D.D. (2002) Assessing the roles of patch quality, area, and isolation in predicting metapopulation dynamics. *Conservation Biology*, **16**, 706–716.

Fountain, T., Husby, A., Nonaka, E., DiLeo, M.F., Korhonen, J.H., Rastas, P., Schulz, T., Saastamoinen, M. & Hanski, I. (2018) Inferring dispersal across a fragmented landscape using reconstructed families in the Glanville fritillary butterfly. *Evolutionary Applications*, **11**, 287–297.

Franzén, M. & Nilsson, S.G. (2010) Both population size and patch quality affect local extinctions and colonizations. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 79–85.

Gilarranz, L.J. & Bascompte, J. (2012) Spatial network structure and metapopulation persistence. *Journal of Theoretical Biology*, **297**, 11–16.

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

Grilli, J., Barabás, G. & Allesina, S. (2015) Metapopulation persistence in random fragmented landscapes. *PLoS computational biology*, **11**, e1004251.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology letters*, **8**, 993–1009.

Hambäck, P.A. & Englund, G. (2005) Patch area, population density and the scaling of migration rates: the resource concentration hypothesis revisited. *Ecology letters*, **8**, 1057–1065.

Hamel, S., Killengreen, S., Henden, J.A., Yoccoz, N. & Ims, R. (2013) Disentangling the importance of interspecific competition, food availability, and habitat in species occupancy: recolonization of the endangered Fennoscandian arctic fox. *Biological conservation*, **160**, 114–120.

Hanski, I. (1991) Single-species metapopulation dynamics: concepts, models and observations. *Metapopulation dynamics: empirical and theoretical investigations*, pp. 17–38. Elsevier.

Hanski, I. (1994a) Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology & Evolution*, **9**, 131–135.

Hanski, I. (1994b) A practical model of metapopulation dynamics. *Journal of Animal Ecology*, pp. 151–162.

Hanski, I. (1999a) Habitat connectivity, habitat continuity, and metapopulations in dynamic landscapes. *Oikos*, pp. 209–219.

Hanski, I. (1999b) *Metapopulation Ecology*, vol. ISSN 1746-3130. Oxford Series in Ecology and Evolution, Oxford, Egnland.

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

Hanski, I. & Ovaskainen, O. (2000) The metapopulation capacity of a fragmented landscape. *Nature*, **404**, 755.

Hanski, I., Schulz, T., Wong, S.C., Ahola, V., Ruokolainen, A. & Ojanen, S.P. (2017) Ecological and genetic basis of metapopulation persistence of the

Glanville fritillary butterfly in fragmented landscapes. *Nature communications*, **8**, 14504.

Hanski, I. & Simberloff, D. (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. *Metapopulation biology*, pp. 5–26. Elsevier.

Hanski, I.A. (2011) Eco-evolutionary spatial dynamics in the Glanville fritillary butterfly. *Proceedings of the National Academy of Sciences*, **108**, 14397–14404.

Hill, J., Thomas, C. & Lewis, O. (1996) Effects of habitat patch size and isolation on dispersal by *Hesperia comma* butterflies: implications for metapopulation structure. *Journal of animal ecology*, pp. 725–735.

Hodgson, J.A., Moilanen, A. & Thomas, C.D. (2009) Metapopulation responses to patch connectivity and quality are masked by successional habitat dynamics. *Ecology*, **90**, 1608–1619.

Howe, R.W., Davis, G.J. & Mosca, V. (1991) The demographic significance of "sink" populations. *Biological Conservation*, **57**, 239–255.

Howell, P.E., Muths, E., Hossack, B.R., Sigafus, B.H. & Chandler, R.B. (2018) Increasing connectivity between metapopulation ecology and landscape ecology. *Ecology*, **99**, 1119–1128.

Ims, R.A. & Andreassen, H.P. (2005) Density-dependent dispersal and spatial population dynamics. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 913–918.

Ims, R.A., Petter Leinaas, H. & Coulson, S. (2004) Spatial and temporal variation in patch occupancy and population density in a model system of an arctic Collembola species assemblage. *Oikos*, **105**, 89–100.

Kleinberg, J.M. (1999) Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, **46**, 604–632.

Kuussaari, M., Nieminen, M. & Hanski, I. (1996) An experimental study of migration in the Glanville fritillary butterfly Melitaea cinxia. *Journal of animal Ecology*, pp. 791–801.

Kuussaari, M., Van Nouhuys, S., Hellmann, J.J. & Singer, M.C. (2004) Larval biology of checkerspots. *On the wings of checkerspots: a model system for population biology. Oxford University Press, Oxford*, pp. 138–160.

Laine, A.L. (2004) A powdery mildew infection on a shared host plant affects the dynamics of the Glanville fritillary butterfly populations. *Oikos*, **107**, 329–337.

Lamy, T., Pointier, J.P., Jarne, P. & David, P. (2012) Testing metapopulation dynamics using genetic, demographic and ecological data. *Molecular Ecology*, **21**, 1394–1410.

Lande, R. (1988) Demographic models of the northern spotted owl (Strix occidentalis caurina). *Oecologia*, **75**, 601–607.

Lipcius, R.N., Eggleston, D.B., Schreiber, S.J., Seitz, R.D., Shen, J., Sisson, M., Stockhausen, W.T. & Wang, H.V. (2008) Importance of metapopulation connectivity to restocking and restoration of marine species. *Reviews in Fisheries Science*, **16**, 101–110.

MacArthur, R.H. (1984) *Geographical ecology: patterns in the distribution of species.* Princeton University Press.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L. & Hines, J.E. (2017) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence.* Elsevier.

Menéndez, R. & Thomas, C.D. (2000) Metapopulation structure depends on spatial scale in the host-specific moth Wheeleria spilodactylus (Lepidoptera: Pterophoridae). *Journal of Animal Ecology*, **69**, 935–951.

Minor, E.S. & Urban, D.L. (2007) Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecological applications*, **17**, 1771–1782.

Mortelliti, A., Amori, G. & Boitani, L. (2010) The role of habitat quality in fragmented landscapes: a conceptual overview and prospectus for future research. *Oecologia*, **163**, 535–547.

Newman, M.E. (2003) The structure and function of complex networks. *SIAM review*, **45**, 167–256.

Ojanen, S.P., Nieminen, M., Meyke, E., Pöyry, J. & Hanski, I. (2013) Long-term metapopulation study of the Glanville fritillary butterfly (Melitaea cinxia): survey methods, data management, and long-term population trends. *Ecology and Evolution*, **3**, 3713–3737.

with contributions from others, G.R. (2017) *gbm: Generalized Boosted Regression Models*. R package version 2.1.3.
URL `https://CRAN.R-project.org/package=gbm`

Ovaskainen, O. (2003) Habitat destruction, habitat restoration and eigenvector–eigenvalue relations. *Mathematical biosciences*, **181**, 165–176.

Ovaskainen, O. (2017) The interplay between immigration and local population dynamics in metapopulations. *Annales Zoologici Fennici*, vol. 54, pp. 113–121. BioOne.

Ovaskainen, O. & Hanski, I. (2003) How much does an individual habitat fragment contribute to metapopulation dynamics and persistence? *Theoretical Population Biology*, **64**, 481–495.

Ovaskainen, O. & Hanski, I. (2004) From individual behavior to metapopulation dynamics: unifying the patchy population and classic metapopulation models. *The American Naturalist*, **164**, 364–377.

Ovaskainen, O. & Saastamoinen, M. (2018) Frontiers in Metapopulation Biology: The Legacy of Ilkka Hanski. *Annual Review of Ecology, Evolution, and Systematics.*

Prugh, L.R., Hodges, K.E., Sinclair, A.R. & Brashares, J.S. (2008) Effect of habitat area and isolation on fragmented animal populations. *Proceedings of the National Academy of Sciences*, **105**, 20770–20775.

Rabasa, S.G., Gutiérrez, D. & Escudero, A. (2008) Relative importance of host plant patch geometry and habitat quality on the patterns of occupancy, extinction and density of the monophagous butterfly Iolana iolas. *Oecologia*, **156**, 491–503.

Ricketts, T.H. (2001) The matrix matters: effective isolation in fragmented landscapes. *The American Naturalist*, **158**, 87–99.

Rosa, E., Woestmann, L., Biere, A. & Saastamoinen, M. (2018) A plant pathogen modulates the effects of secondary metabolites on the performance and immune function of an insect herbivore. *Oikos.*

Rozenfeld, A.F., Arnaud-Haond, S., Hernández-García, E., Eguíluz, V.M., Serrão, E.A. & Duarte, C.M. (2008) Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences*, pp. pnas–0805571105.

Smith, A.T. & Peacock, M.M. (1990) Conspecific attraction and the determination of metapopulation colonization rates. *Conservation Biology*, **4**, 320–323.

Steffan-Dewenter, I. & Schiele, S. (2008) Do resources or natural enemies drive bee population dynamics in fragmented habitats? *Ecology*, **89**, 1375–1387.

Thomas, C. (1994) Extinction, colonization, and metapopulations: environmental tracking by rare species. *Conservation Biology*, **8**, 373–378.

Thomas, C. & Harrison, S. (1992) Spatial dynamics of a patchily distributed butterfly species. *Journal of Animal Ecology*, pp. 437–446.

Thomas, J., Bourn, N., Clarke, R., Stewart, K., Simcox, D., Pearman, G., Curtis, R. & Goodger, B. (2001) The quality and isolation of habitat patches both determine where butterflies persist in fragmented landscapes. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 1791–1796.

Urban, D. & Keitt, T. (2001) Landscape connectivity: a graph-theoretic perspective. *Ecology*, **82**, 1205–1218.

Urban, D.L., Minor, E.S., Treml, E.A. & Schick, R.S. (2009) Graph models of habitat mosaics. *Ecology letters*, **12**, 260–273.

Zamborain-Mason, J., Russ, G.R., Abesamis, R.A., Bucol, A.A. & Connolly, S.R. (2017) Network theory and metapopulation persistence: incorporating node self-connections. *Ecology Letters*, **20**, 815–831.

Tables

Table 1: The potential directionality ('Sign') of each covariate group ('patch area', 'habitat', 'spatial', and 'network') on metapopulation dynamics. Metapopulation dynamics may be positively (↑) or negatively (↓) associated to the covariate group. The relationship between covariate group and metapopulation process (occupancy, colonization, or extinction) may be unclear or could be either positive or negative (↑/↓). Lastly, the putative explanation for the relationship is provided in the 'Directionality' column.

Response	Model	Sign	Directionality
Occupancy			
	Patch area	↑	Larger patches support larger populations
	Habitat	↑/↓	Mildew and grazing reduce occupancy, resources increase occupancy
	Spatial	↑/↓	Spatial patterns in historical introductions drive occupancy
	Network	↑	Central patches are more likely to be occupied
Colonization			
	Patch area	↑	Larger patches are bigger colonization targets
	Habitat	↑/↓	Mildew and grazing reduce colonization, resources increase colonization
	Spatial	↑/↓	Spatial patterns in historical introductions drive colonization
	Network	↑	Central patches are more likely to receive immigrants
Extinction			
	Patch area	↓	Larger patches have less extinction prone populations
	Habitat	↑/↓	Mildew and grazing enhance extinction, resources decrease extinction
	Spatial	↑/↓	Spatial patterns of occupancy influence extinction
	Network	↓	Central patches are less likely to go extinct

Table 2: The identities of each of the covariates included in the submodels (e.g., habitat). All covariates were included in the full model, in order to estimate overall importance of each covariate. The measurement or estimation of each variable is described in more detail in the *Variables influencing occupancy and colonization* Methods section.

Group	Variable	Description
Patch area	log(Patch area)	Area of habitat patch in km^2
Habitat	Resource availability	Total resources on ordinal scale (0-6)
	Plantago lanceolata	Plantago resources on ordinal scale (0-3)
	Veronica spicata	Veronica resources on ordinal scale (0-3)
	Grazing pressure	Estimated percentage of plants grazed
	Mildew infection	Fraction of time mildew pathogen found in given patch
Spatial	Latitude	Latitudinal coordinate of patch (decimal degrees)
	Longitude	Longitudinal coordinate of patch (decimal degrees)
Network	Betweenness centrality	Patch importance measure focused on *stepping stones*
	Closeness centrality	Importance measure based on the entire dispersal network
	Degree centrality	Local-scale importance of dispersal connections
	Eigenvector centrality	Importance estimated by connections to important patches

709 # Figures

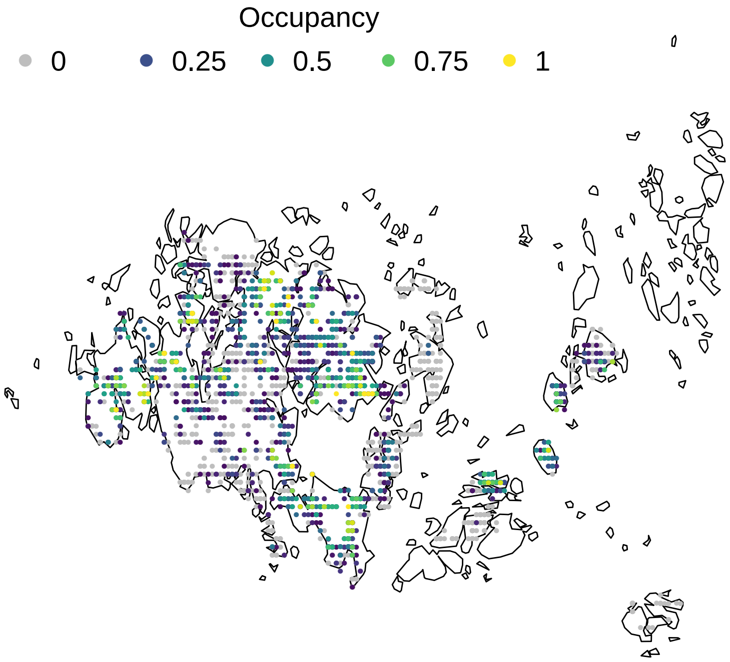

Figure 2: Maps of the Åland islands showing the distribution of sampled habitat patches as part of the monitoring effort, with habitat patches colored by the fraction of times the sampled patch was occupied between the period of 2000-2017. Patches in grey are those in which Glanville fritillary butterfly (*M. cinxia*) was never recorded.

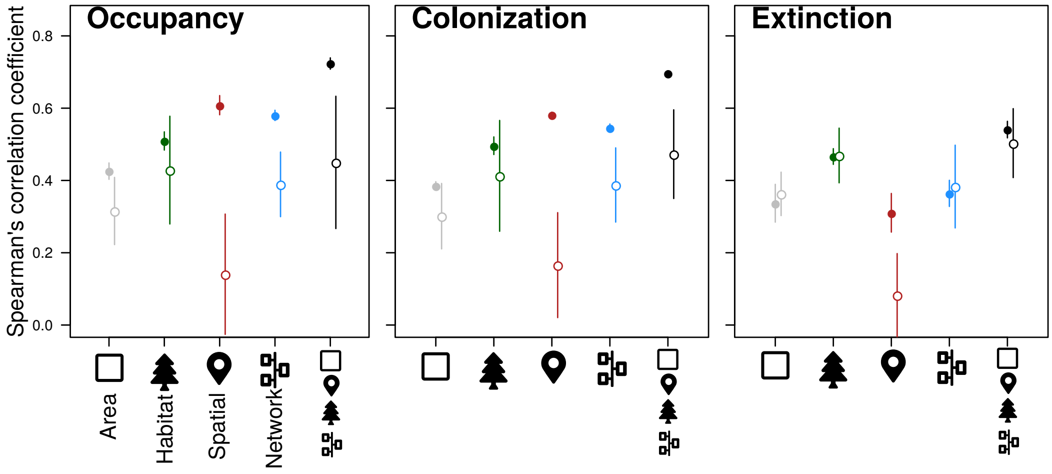

Figure 3: Model performance – defined as Spearman's correlation coefficient between model-predicted values and empirical data from a subset of data not used to train the model – for each of the candidate models with both random cross-validation (closed circles) and spatially-stratified cross-validation (open circles). Plotted points correspond to average correlations across the ten cross-validated models, and bars correspond to standard deviation. Glyphs are from Font Awesome (`https://fontawesome.com/`).

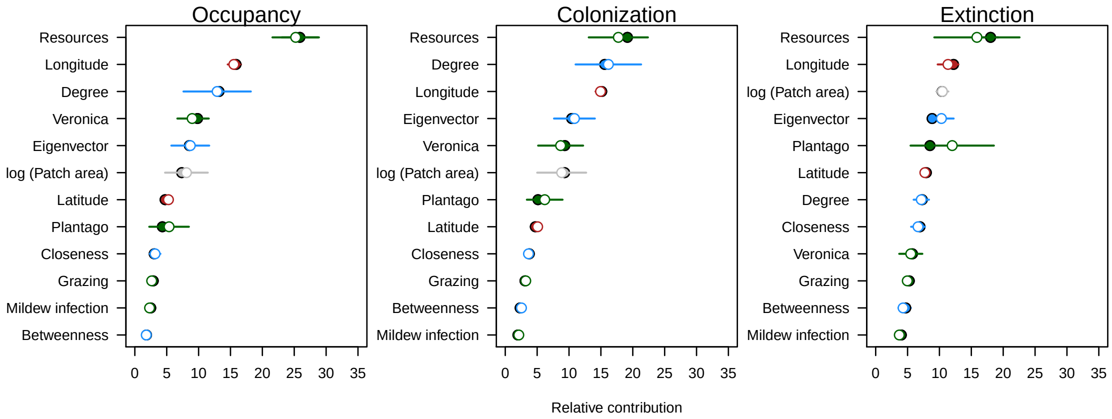

Figure 4: The trained boosted regression tree models revealed that resource availability, degree centrality, and longitude were important predictors of patch occupancy, colonization, and extinction. Variable relative importance remains quite similar with both random cross-validation (closed circles) and spatially-stratified cross-validation (open circles). Bars represent standard deviation across the set of five trained models on different subsets of data.

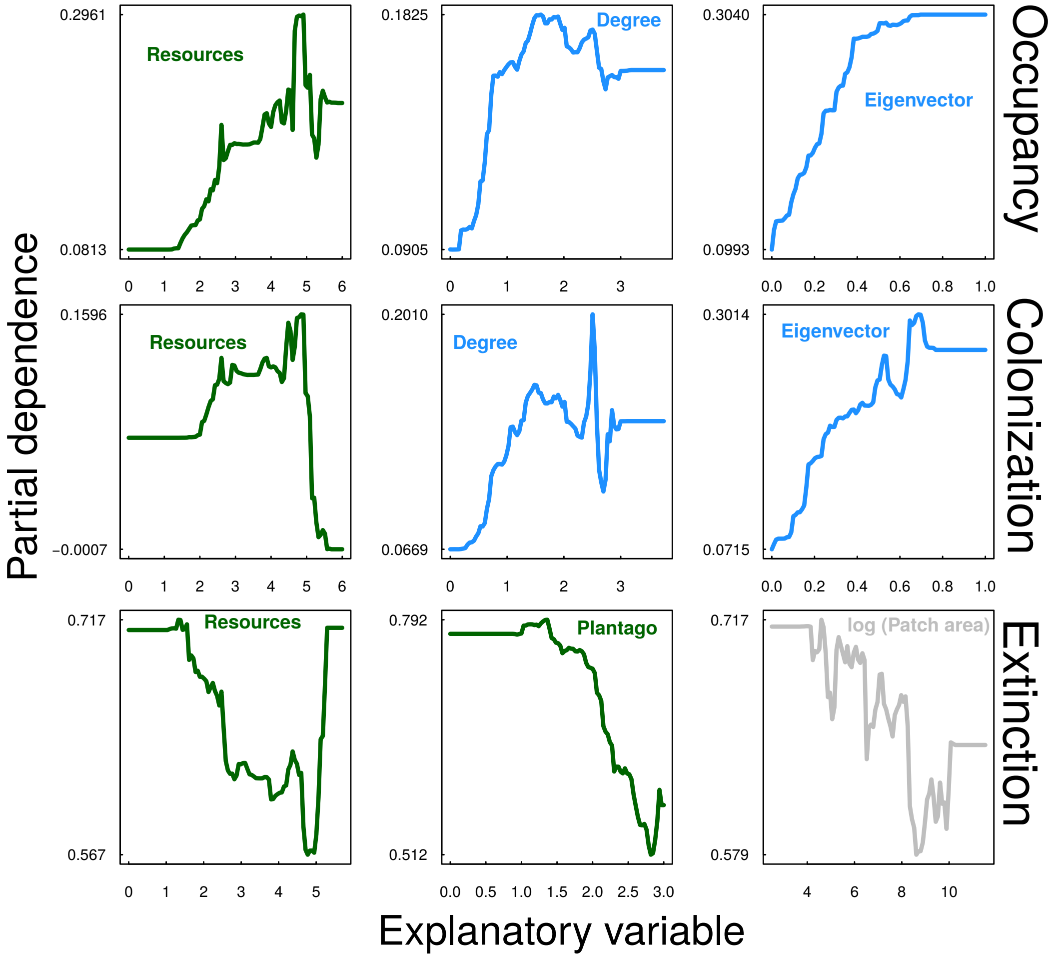

Figure 5: Partial dependence plots for the top three predictors in the boosted regression tree model of *M. cinxia* occupancy (top row), colonization (middle row), and extinction (bottom row), showing the relationships between each metapopulation process and the top three predictive variables in each model when models were cross validated by spatially stratification. The most important variables in the full models of occupancy, colonization, and extinction tended to be related to resource availability and connectivity.