# ECOGRAPHY

## Research

# Long-term demographic surveys reveal a consistent relationship between average occupancy and abundance within local populations of a butterfly metapopulation

Torsti Schulz, Jarno Vanhatalo and Marjo Saastamoinen

T. Schulz (https://orcid.org/0000-0001-8940-9483) ✉ (torsti.schulz@helsinki.fi), J. Vanhatalo (https://orcid.org/0000-0002-6831-0211) and M. Saastamoinen (https://orcid.org/0000-0001-7009-2527), Organismal and Evolutionary Biology Research Programme, Univ. of Helsinki, Finland. JV also at: Dept of Mathematics and Statistics, Univ. of Helsinki, Finland. MA also at: Helsinki Inst. of Life Science (HiLIFE), Univ. of Helsinki, Finland.

Species distribution models are the tool of choice for large-scale population monitoring, environmental association studies and predictions of range shifts under future environmental conditions. Available data and familiarity of the tools rather than the underlying population dynamics often dictate the choice of specific method – especially for the case of presence–absence data. Yet, for predictive purposes, the relationship between occupancy and abundance embodied in the models should reflect the actual population dynamics of the modelled species. To understand the relationship of occupancy and abundance in a heterogeneous landscape at the scale of local populations, we built a spatio-temporal regression model of populations of the Glanville fritillary butterfly *Melitaea cinxia* in a Baltic Sea archipelago. Our data comprised nineteen years of habitat surveys and snapshot data of land use in the region. We used variance partitioning to quantify relative contributions of land use, habitat quality and metapopulation covariates. The model revealed a consistent and positive, but noisy relationship between average occupancy and mean abundance in local populations. Patterns of abundance were highly variable across years, with large uncorrelated random variation and strong local population stochasticity. In contrast, the spatio-temporal random effect, habitat quality, population connectivity and patch size explained variation in occupancy, vindicating metapopulation theory as the basis for modelling occupancy patterns in fragmented landscapes. Previous abundance was an important predictor in the occupancy model, which points to a spillover of abundance into occupancy dynamics. While occupancy models can successfully model large-scale population structure and average occupancy, extinction probability estimates for local populations derived from occupancy-only models are overconfident, as extinction risk is dependent on actual, not average, abundance.

Keywords: analysis of variance, habitat quality, landscape ecology, metapopulation ecology, occupancy–abundance relationship, species distribution modeling

NORDIC SOCIETY OIKOS

www.ecography.org

## Introduction

Observed declines in insect abundance around the world remind us of the importance of understanding changes in species' abundance in addition to looking at range shifts and population extinctions (Conrad et al. 2006, Shortall et al. 2009, Cameron et al. 2011, Koh et al. 2016, Hallmann et al. 2017). At the same time, deficiencies in some studies highlight the challenges in analyzing and interpreting population dynamics of as diverse and dynamic a group as insects (Georgina et al. 2015, 2016, Komonen et al. 2019, Sánchez-Bayo and Wyckhuys 2019, Thomas et al. 2019).

Occupancy and abundance patterns both arise from spatio-temporal variation in growth rates and dispersal. Notably, ecological theory does not assert that distribution, occupancy and abundance respond similarly to variation in environmental conditions, as they are different functions of those population dynamic processes. Spatio-temporal variation in the environment leads not only to variation in the dispersal and local growth rates, but the relationship between them. Especially where density-dependence affects growth and dispersal, this could lead to different relationships between occupancy, distribution and abundance. While ecological studies demonstrate that these population measures are often correlated (Gaston et al. 2000, Cowley et al. 2001), this does not imply that their relationship is fixed across different contexts and taxa (Holt et al. 2002) or driven by the same processes. Suggested mechanisms that give rise to different, mostly positive, occupancy–abundance relationships both within and among species include for example sampling processes, metapopulation dynamics, species-specific environmental responses and spatial variation in habitat quality (Hanski 1991, Hanski and Gyllenberg 1997, Freckleton et al. 2005, 2006).

In applied spatial ecology, the availability of methods and data often limit us to analyses not derived from assumptions about species' demography. Consequentially, these methods often introduce implicit relationships between occupancy and abundance into our analyses. Species distributions and occupancy models usually use only occupancy data and model patterns instead of demographic processes (Pearce and Ferrier 2001, Sileshi 2007, Keith et al. 2008, Duff et al. 2012). Although species distribution models built on abundance data give a more detailed view of the relationship between population density and the environment (Kallasvuo et al. 2017), they still implicitly assume a simple relationship between patterns of occupancy and abundance (Sileshi et al. 2009, Dallas and Hastings 2018). Models of abundance conditioned on covariates that use for example a Poisson or negative binomial distribution yield occupancy probabilities as a simple function of the distributional form (Holt et al. 2002). In environments where dispersal is not a limiting factor and there is no strong spatial correlation in habitat quality, such simple relationships can be reasonable approximations, but in fragmented landscapes where occupancy is a function of both colonizations and local population dynamics, we should not let this assumption stand untested. Even models grounded directly in population ecology – such as the stochastic patch occupancy models of metapopulation ecology – make simplifications that limit the ability of models to distinguish how environmental variation or interspecific interactions can lead to diverging patterns of occupancy and abundance (Keeling 2002, Etienne et al. 2004). These models, which are fit to records of presence and absence of the study species, often assume a simple relationship between carrying capacity and patch characteristic, such as area and habitat quality, and use this relationship as the basis of extinction probabilities. More principled patch occupancy models derive a surrogate of population size from an underlying individual based model, but ignore stochastic variation in that size (Ovaskainen and Hanski 2004b). In both cases, if the model includes environmental effects on occupancy they also implicitly affect abundance in a predetermined fashion, unless the model distinguishes between environmental effects on dispersal and local reproduction (Harrison et al. 2011). Such deficiencies are most critical when predicting under novel environmental conditions and outside a species' current range. An extreme example are bioclimatic envelope models lacking any population dynamic component that are used to predict range shifts under climate change (Araújo et al. 2005, Lewthwaite et al. 2018), in spite of the availability dynamic alternatives (Keith et al. 2008, Buckley et al. 2010, Leroux et al. 2013).

Here, we apply joint occupancy–abundance modelling to the long-term demographic survey of the Glanville fritillary in the Åland Islands (Ojanen et al. 2013). We assess how occupancy and abundance respond to environmental heterogeneity in this large metapopulation. The data combine reliable demographic surveys of the number of larval groups in local populations and details on habitat quality with a comparatively large spatial extent suitable for understanding the relationship between occupancy and abundance in a heterogeneous landscape. We study how the spatial structure of habitat, population dynamics, habitat quality and the composition of the landscape contribute to occupancy and abundance and how these contributions differ at the level of local populations in the metapopulation. The choice of study system governs our expectations. Given the host–plant specificity of the butterfly's larval stages and the classic metapopulation structure of the populations, we expect to see large effects of variation in habitat quality, patch size and the connectivity of the habitat. We expect less pronounced effects of land use as effects of land use are mediated by direct measures of habitat quality. Still, land use is expected to have a more direct role through effects on dispersal (DiLeo et al. 2018). Comparing occupancy and abundance, we expect that occupancy be largely determined by landscape structure, that is, the distribution of habitat and land use, while local variation in the habitat, especially in terms of area and quality would be more important for abundance.

We use a hurdle model with a spatio-temporal random effect to account for potentially confounding unobserved variation and spatio-temporal correlation between our observations. The model captures some elements of population

dynamics and is simple enough to fit large data sets. The random effects in the model help quantify how unexplained variation is structured and reflect a lower bound on the uncertainty that our model would carry over when extrapolating into different environments. As the study is observational, we cannot interpret covariate effects causally. Instead, we use variance partitioning to quantify which factors matter in our study system and how much we can potentially expect to explain in other systems. We batch our covariates into classes that group together related covariates in terms of what they represent and where they are measured. This allows us to draw more general conclusions than from individual covariates whose effects are more likely species and site dependent.

## Material and methods

### Study system

We study the metapopulations of the Glanville fritillary butterfly (*Melitaea cinxia* Linnaeus 1758, family Nymphalidae) in the Åland Islands in the Baltic Sea. At this northern range margin the butterfly has a single generation per year with adults emerging in late May or early June and the flight season lasting for approximately one month (Hanski et al. 1994, Nieminen et al. 2004, Kahilainen et al. 2018). The larvae live in family groups and overwinter in conspicuous tent-like silken nests that they build late in the summer or in early fall (Hanski et al. 1995, Kuussaari and Singer 2017). Though endangered in Finland, the species is not under strict protection (Hyvärinen et al. 2019).

In the Åland Islands, the butterfly inhabits networks of dry meadows, pastures, roadsides and rocky outcrops where one or both of its larval host plants, ribwort plantain *Plantago lanceolata* and spiked speedwell *Veronica spicata*, grow (Hanski et al. 1994). The study area, surveyed annually since the 1993, covers about $50 \times 70 \, \text{km}^2$ with over four thousand potential habitat patches known to date. The butterfly occupies annually between 1 and 30 percent of the habitat patches with annual turnover usually comprising between 50 and 200 extinctions and colonizations (Hanski et al. 1995, Ojanen et al. 2013). The system has become a model for study of metapopulation dynamics (Hanski et al. 1995, Hanski 1999, Nieminen et al. 2004, Ovaskainen and Saastamoinen 2018).

Every September surveyors visit most of the habitat patches to measure habitat quality and count the number of the butterfly's larval nests (Hanski et al. 1995, Ojanen et al. 2013). The probability of detecting at least one nest in an occupied habitat patch is 70–90 percent (Hanski et al. 1995, 2017, Nieminen et al. 2004, Ojanen et al. 2013). As intensive resurveys for undiscovered habitat in 1998 and 1999 found over two thousand new habitat patches (Ojanen et al. 2013), we restrict our study to the more complete data collected since year 2000. The data comprise 62 659 patch observations spanning a 19-yr period. The records include 9780 observations of occupied patches.

### Land use data

We obtained land use data from the National Land Survey of Finland's (NLS) Topographic database (acquired 08/2017). They represent a recent snapshot of the landscape state. We restricted ourselves to land use and land cover data represented by the categories Terrain/1, Terrain/2 and Road network in the Topographic database. We simplified the land use classification to reduce the number of land use covariates. The resulting data include agricultural fields, horticulture, meadows, marshes and bogs, open water bodies, built and recreational areas, and roads (Supplementary material Appendix 1 Table A1). Over half of the agricultural fields in the region produce hay for ensilage and pasture. Other major field crops include various cereals, while apple orchards dominate horticultural land use. Additionally, we include data on forested areas from the 2011 Multi-source National Forest Inventory provided by the National Resources Inst. Finland. We excluded open bogs from our forest class as they are either treeless or only sparsely forested (Mäkisara et al. 2016).

### Covariate categorization

To study how different aspects of the landscape affect occupancy and abundance of the butterfly, we group our covariates into four categories: population (P), metapopulation (M), habitat quality (H) and land use (L). We also divide our covariates in terms of their spatial scope, that is, whether they are measurements from within the habitat patch or from the surrounding landscape (Fig. 1).

The local population covariates (P) include population size (number of larval nests) and occupancy status in the previous year. The metapopulation covariates (M), habitat patch area and population connectivity, are considered the primary predictors of occupancy in metapopulation ecology (Hanski 1999). We use a model based connectivity measure, which comprises the individual immigrant contributions from all other patches based on their spatial configuration, area and population size in the previous year (Hanski 1994, Supplementary material Appendix 2).

The habitat quality covariates (H) include an ordinal measure of the abundance of either of the host plant species, a binary indicator for the simultaneous presence of both host plants, the proportion of desiccated host plants, the proportion of grazed vegetation in the habitat patch, and an indication of the presence of powdery mildew *Podosphaera plantaginis* on the host plant *Plantago lanceolata* (Ojanen et al 2013). We simplified the habitat quality measures for the analyses (Supplementary material Appendix 3).

The land use covariates (L) include three measures for each land use class: 1) proportion in the habitat patch, 2) proportion outside the patch edge in a 10-m buffer and 3) distance weighed proportion in a large buffer around the habitat patches (Aue et al. 2012, Chandler and Hepinstall-Cymerman 2016, Miguet et al. 2017). Land use within the patch can capture aspects of habitat quality not recorded
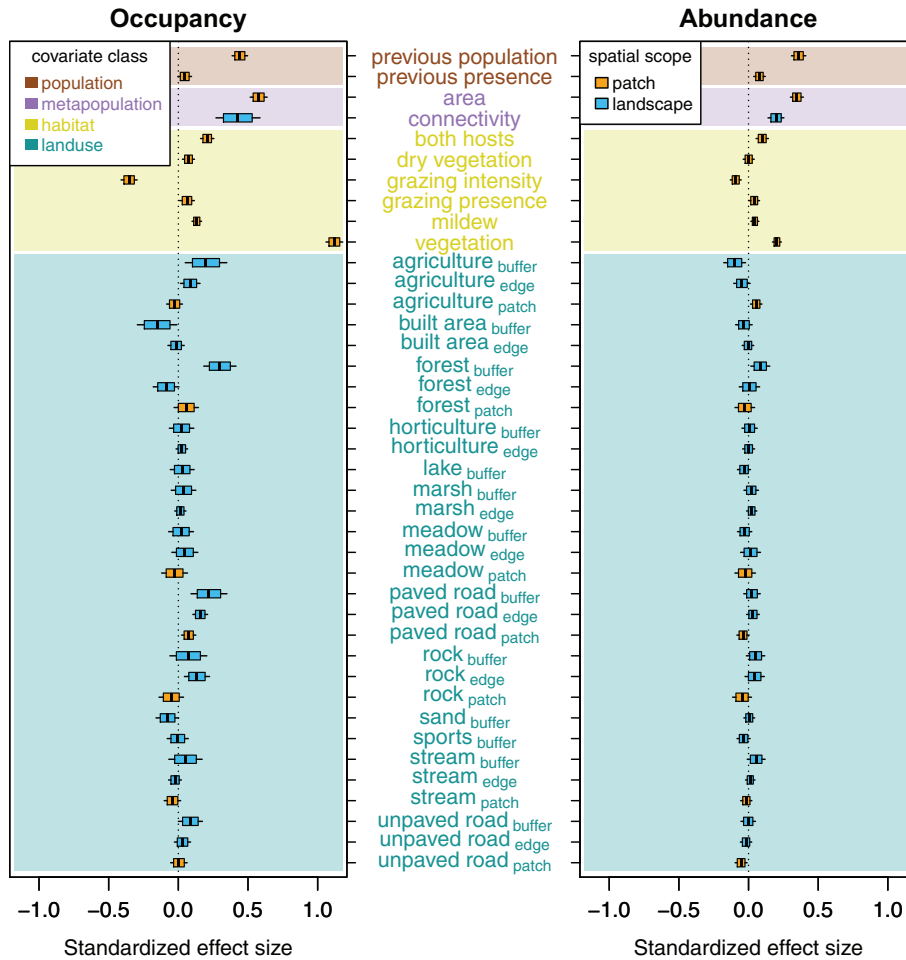
Figure 1. Posterior means of the standardized effect sizes of the occupancy and abundance model's covariates. The boxes show the 80% credible intervals and the whiskers extend to the 95% credible interval. Blue depicts effects that pertain to the habitat patch and orange depicts landscape-based effects. The label and background colours depict the covariate classes.

during the surveys, while the land use at the edge of the patch can have direct effects on habitat microclimate, for example through shading, and also impede or facilitate immigration and emigration of adult butterflies. Land use in the wider surrounding area can capture effects of landscape composition on long-distance dispersal or larger microclimatic conditions. The scales of the environmental effects may vary among the covariates (Martin and Fahrig 2012), but given the large number of covariates included in our study we opted for simplicity and used the same fixed parameters for the distance weighing function for all land use classes (Supplementary material Appendix 4). Of the resulting landscape composition measures, originally three per land-use class, we kept only those that occurred in at least one percent of all observations.

**Model**

We used a hurdle model to separate the effect of environmental variation on butterfly occupancy and abundance (Welsh et al. 1996). Due to the fast turnover in the

metapopulation and variation in habitat patch isolation, absences are not necessarily indicators of unsuitable habitat. Thus, we exploit the structure of the hurdle model to distinguish processes that govern occupancy and abundance. The occupancy component mimics the colonization–extinction dynamics of the metapopulation while the abundance component should reflect the effects of site characteristics on local reproductive performance.

We apply a generalized linear modelling approach to model occupancy and abundance as a linear combination of covariates, independent yearly and per-patch random effects, and a spatiotemporal random effect. We model abundance $y$ with a negative binomial distribution and occupancy with a Bernoulli distribution. The model is then

$$p(y_i) = \begin{cases} 1 - \theta_i & y_i = 0 \\ \theta_i \times \dfrac{\text{Negative binomial}(y_i | \lambda_i, n)}{1 - \text{Negative Binomial CDF}(0 | \lambda_i, n)} & y_i \geq 1 \end{cases} \quad (1)$$

where

$$\text{logit}(\theta_i) = X_i\beta_o + a_o(i) + u_o(t_i) + z_o(s_i, t_i) \qquad (2)$$

$$\log(\lambda_i) = X_i\beta_a + a_a(i) + u_a(t_i) + z_a(s_i, t_i) \qquad (3)$$

and $y_i$ is the abundance measured as the observed nest count, $\theta$ is the probability of presence, $\lambda$ the expected butterfly abundance in an occupied patch, $X$ is a matrix with covariates in its columns, $\beta_o$ and $\beta_a$ are mutually independent column vectors with covariate weights, $a_o$ and $a_a$ are mutually independent per-patch iid random effects, $u_o$ and $u_a$ are mutually independent yearly iid random effects, $z_o$ and $z_a$ are mutually independent spatio-temporal random effects, $s$ is a vector of observation coordinates, $t$ is a vector of observation years, $n$ is the overdispersion parameter. The random effects represent unmeasured variation at the level of patches ($a$), unmeasured large-scale variability between the years ($u$) and unmeasured spatio-temporally correlated variation ($z$); the overdispersion parameter $n$ can also be interpreted as a per-observation random effect representing, for example, unmeasured environmental stochasticity (Lindén and Mäntyniemi 2011). The use of the above hurdle model, where the likelihood terms for zero and greater than zero observations are separable, is justified by the high rate of detection in the survey, which ensures a low proportion of false absences in the data and minimizes the effects of potential variation in detection rates (Bried and Pellet 2012).

We gave independent $\mathcal{N}(0, \sqrt{10})$ priors for all the linear weights except for the intercept, which had a $\mathcal{N}(0, \sqrt{100})$ prior. All covariates were normalized to have standard deviations of one and a mean of zero. Before normalization, we log transformed patch area and connectivity and $\log(x+1)$ transformed host plant abundance (vegetation), and previous population size. We implemented the model using R-INLA (ver. 18.07.12, Supplementary material Appendix 5) which provides efficient approximate computation for a wide class of latent Gaussian models (Rue et al. 2009, Lindgren et al. 2011, Lindgren and Rue 2015).

### Partitioning of variation

To quantify the contribution of different measures of landscape heterogeneity to butterfly occupancy and abundance, we examine how much the covariates in each covariate category (P, M, H and L) and the random effects contribute to the total posterior variation in the linear predictors $\text{logit}(\theta)$ and $\log(\lambda)$. This corresponds to the 'finite-population' view of Bayesian analysis of variance (Gelman et al. 2014). In the spatio-temporal modelling context, Yuan et al. (2017) used this method for exploratory model checking, while Ovaskainen et al. (2017) applied it in hierarchical models in community ecology. As our study system consists of a discrete set of populations, we calculate the measure over

all observation units (Supplementary material Appendix 6). To understand how the different components of the model contribute to explaining variation within and among patches, we also partition the observations by patch and apply the law of total (co)variance to calculate within- and among-patch variances for all components over all patches (Table 2).

### Results

#### Covariate effects

Overall, the population, metapopulation and habitat quality covariates had clear effects on occupancy and abundance (Fig. 1, Supplementary material Appendix 6 Table A2). Of the population covariates, the size of the population in the previous year increased occupancy probabilities; previous occupancy status alone had no discernible effect. The odds of occupancy when comparing a patch with ten larval nests to a patch with only one nest in the previous year are over two and half times higher (Fig. 2). While both size and occupancy of the population in the previous year increased abundance, population size in the previous year had clearly a stronger effect and together with habitat patch area it was the covariate
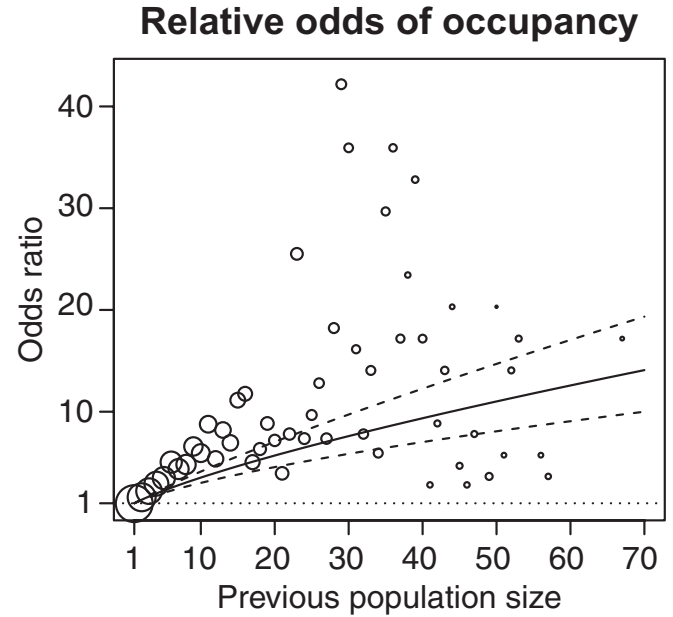


**Relative odds of occupancy**

Figure 2. Estimated ratio of odds of occupancy for populations with a single nest the year before compared to odds of occupancy for other population sizes. The posterior mean is marked with a solid line and the 95% credible interval with dashed lines. The circles show observed odds ratios with circle radii proportional to the log of the number of observations with that population size in the previous year $\left(\text{OR}_{\text{obs}}(x) = \dfrac{(n_{y_{t-1}=x, y_t > 0} + 0.5)/(n_{y_{t-1}=x, y_t = 0} + 0.5)}{(n_{y_{t-1}=1, y_t > 0} + 0.5)/(n_{y_{t-1}=1, y_t = 0} + 0.5)}\right)$. Population sizes represented by less than four observations in the numerator are excluded.

with the strongest effect. Habitat patch area had a larger effect than population connectivity for both occupancy and abundance (Fig. 1). Of the habitat quality covariates, host plant abundance (vegetation in Fig. 1) had the strongest effect, and for the occupancy model, it was overall the most important covariate. Simultaneous presence of both host plant species (both hosts in Fig. 1) was additionally associated with higher abundance and probability of occupancy. A higher proportion of dry or desiccated host plant vegetation, as assessed during the fall survey, was associated with slightly higher occupancy probabilities. Similarly, the presence of a powdery mildew infection on the host plant *Plantago lanceolata* increased occupancy and abundance slightly. Intense grazing of habitat predicted a drastically reduced occupancy probability and slightly lower abundance. In contrast, the presence of grazing itself slightly increased occupancy and abundance.

Land use covariates had mostly minor effects, except for the effects on occupancy of land use outside of the habitat patches (Fig. 1). Forests and agricultural land in the surrounding areas, and agricultural land and rocky areas along the edges of the habitat patches increased patch occupancy, while surrounding built-up areas decreased it. Paved roads increased occupancy probabilities generally as did unpaved roads in the surrounding areas. The butterfly abundances were also slightly higher in patches surrounded by forests. Agricultural land use within the patch increased abundance, while its presence in the surroundings decreased it. Unpaved roads decreased abundance when they passed through the actual habitat patch.

## Variation explained

Most variation in occupancy patterns was explained by the yearly iid and spatio-temporal random effects (posterior mean 0.40, Table 1). In contrast, the patch level random effect had only a minor role. Overall, processes related to the landscape were more important than variation of measures pertaining only to the habitat patches themselves (0.54 versus 0.28). Yet, habitat quality was the most important covariate class in terms of variation explained (0.14). Metapopulation covariates, land use outside of the patches, and population covariates each explained only a few percent of variation, while land use within habitat patches had no discernible effect. The spatio-temporal random effect correlated with population and metapopulation covariates (posterior means 0.33 and 0.35, Supplementary material Appendix 6 Table A3) and the population, metapopulation and habitat quality covariates correlated with each other (0.26–0.40). The spatio-temporal random field correlated strongly with connectivity and the population covariates (posterior mean correlations 0.64 and 0.33, Supplementary material Appendix 8 Table A3, Fig. A2), which had a particularly strong effect on the results: Both explained only two percent of variation in occupancy, but of all the covariates had the highest correlations with the linear predictor logit($\theta$) (squared mean posterior correlation 0.36, Supplementary material Appendix 8 Table A5, Fig. A2) suggesting strong confounding between connectivity, the population covariates and the spatio-temporal random effect in the occupancy model.

Random effects related to the whole landscape accounted for the largest proportion of the variation in abundance (posterior mean 0.41, Table 1), but in contrast to occupancy also patch level random variation was important (0.11). In general, the contributions of patch level and landscape wide processes were almost equal in explaining variation in abundance (0.40 and 0.46). Both population covariates and metapopulation covariates explained a large proportion of variation (0.13 and 0.12). Habitat quality was less important than for occupancy, and just as land use in the landscape, it explained five percent of variation in abundance. The patch and yearly random effects correlated with population covariates (posterior means 0.20 and −0.22, Supplementary material Appendix 6 Table A4), but the spatio-temporal random effect had only low correlations with the other components. The population, metapopulation and habitat quality covariates correlated with each other (posterior means 0.20–0.46).

Table 1. Proportion of variation in the linear predictors (eq. 2–3) explained by their components. The 95% credible intervals are below the mean value in each cell. The totals for each model sum to one when accounting for covariance between the components.

| | Occupancy | | Abundance | |
|---|---|---|---|---|
| | Patch | Landscape | Patch | Landscape |
| Random effects | 0.03 | 0.40 | 0.11 | 0.41 |
| | 0.02–0.04 | 0.36–0.43 | 0.09–0.13 | 0.38–0.44 |
| Population covariates | 0.02 | | 0.13 | |
| | 0.02–0.02 | | 0.11–0.16 | |
| Metapopulation covariates | 0.03 | 0.02 | 0.09 | 0.03 |
| | 0.02–0.04 | 0.01–0.03 | 0.07–0.12 | 0.01–0.05 |
| Habitat quality covariates | 0.14 | | 0.05 | |
| | 0.13–0.15 | | 0.04–0.06 | |
| Land use covariates | 0 | 0.03 | 0.01 | 0.05 |
| | 0–0 | 0.02–0.03 | 0–0.02 | 0.03–0.07 |
| Total | 0.28 | 0.54 | 0.40 | 0.46 |
| | 0.26–0.30 | 0.52–0.57 | 0.37–0.44 | 0.43–0.48 |

Table 2. Proportion of variation within and between patches as explained by the components of the linear predictors (2–3). The 95% credible intervals are below the mean value in each cell. The totals on the final row give posterior estimates of the variation due to within and among patch variability in occupancy and abundance.

| | Occupancy | | Abundance | |
|---|---|---|---|---|
| | Within | Among | Within | Among |
| Random effect: patch | | 0.03 | | 0.11 |
| | | 0.02–0.04 | | 0.09–0.13 |
| Random effect: year | 0.10 | 0.00 | 0.13 | 0.03 |
| | 0.08–0.12 | 0.00–0.00 | 0.11–0.16 | 0.02–0.03 |
| Random effect: spatial | 0.10 | 0.20 | 0.17 | 0.08 |
| | 0.09–0.12 | 0.17–0.23 | 0.15–0.19 | 0.06–0.09 |
| Population covariates | 0.01 | 0.01 | 0.08 | 0.05 |
| | 0.01–0.01 | 0.01–0.01 | 0.06–0.10 | 0.04–0.07 |
| Metapopulation covariates | 0.01 | 0.05 | 0.01 | 0.13 |
| | 0.00–0.01 | 0.04–0.06 | 0.01–0.02 | 0.11–0.16 |
| Habitat quality covariates | 0.07 | 0.07 | 0.02 | 0.03 |
| | 0.07–0.08 | 0.06–0.07 | 0.02–0.03 | 0.02–0.04 |
| Land use covariates | | 0.03 | | 0.04 |
| | | 0.02–0.03 | | 0.03–0.06 |
| Total | 0.33 | 0.67 | 0.38 | 0.62 |
| | 0.31–0.35 | 0.65–0.69 | 0.36–0.40 | 0.60–0.64 |

## Differences between occupancy and abundance

Metapopulation covariates were important for explaining differences among patches, while population and habitat quality covariates contributed almost equally to within- and among-patch variation (Table 2). For within-patch variation in occupancy, the yearly and spatio-temporal random effects and habitat quality explained most variation, while the yearly and spatio-temporal random effects and population covariates explained most within-patch variation in abundance.

Variation in the occupancy and abundance models was higher among patches compared to within patches (posterior means 0.67 and 0.62, Table 2). The spatio-temporal random effect explained more variation in occupancy among patches than within (0.20 and 0.10), while for abundance their relationship was reversed (0.08 and 0.17). This suggests that the random effects explain long-term differences in occupancy over the study region while for abundance there is more yearly varying spatially structured stochasticity. We see the same relationship from the model hyperparameters: the spatio-temporal random field for the occupancy model had higher temporal autocorrelation (posterior mean 0.92, Supplementary material Appendix 9 Table A6) than the model for abundance (0.38) and correlation distance for occupancy (6.1 km) was longer than for abundance (4.4 km). The temporal means, trend and variance of the random effects differed spatially between the occupancy and abundance models (Fig. 3). The occupancy model's spatio-temporal random effect had clear and stable structure, whose mean clearly differentiated areas where the butterfly occurs from unoccupied regions of the landscape. For the abundance model, the temporal variation in the spatio-temporal random effect swamped any permanent large-scale structure (Supplementary material Appendix 10 Fig. A4).

In general, unexplained and random variability both within and especially among patches, was higher for the abundance model (Table 1). The standard deviations of the patch, yearly and spatio-temporal random effects differed less from each other in the abundance model than in the occupancy model, where the spatio-temporal random effect clearly had the highest variance (Supplementary material Appendix 9 Table A6). Together with the lower temporal and spatial autocorrelation, the variance parameters also imply a much higher annual variation of the random effects for the abundance model.

The relationship between log abundance and log odds of occupancy within patches was linear, suggesting an exponential relationship between the odds of occupancy and abundance within patches (Fig. 4). The relationship would deviate from linearity if the effects of the covariates differed strongly between occupancy and abundance or the distribution of covariates in unoccupied habitat was very different.

## Discussion

Our results demonstrate that while similar facets of environmental heterogeneity affect both occupancy and abundance, their relative roles differ in the Glanville fritillary metapopulation. With the notable exception of the strong role of habitat quality, landscape properties affect occupancy more than habitat patch characteristics, while abundance is determined both by local and landscape properties, with habitat quality one local factor among many.

We find very little unstructured variation among patches in the occupancy model: the model covariates and the spatio-temporal random effect explain most variation. Given the strong correlation of the spatio-temporal random field with our connectivity measure and population covariates, we can say that previous year's population size, patch area, habitat quality and connectivity are the main drivers of occupancy, as expected in a classic metapopulation (Hanski 1999).
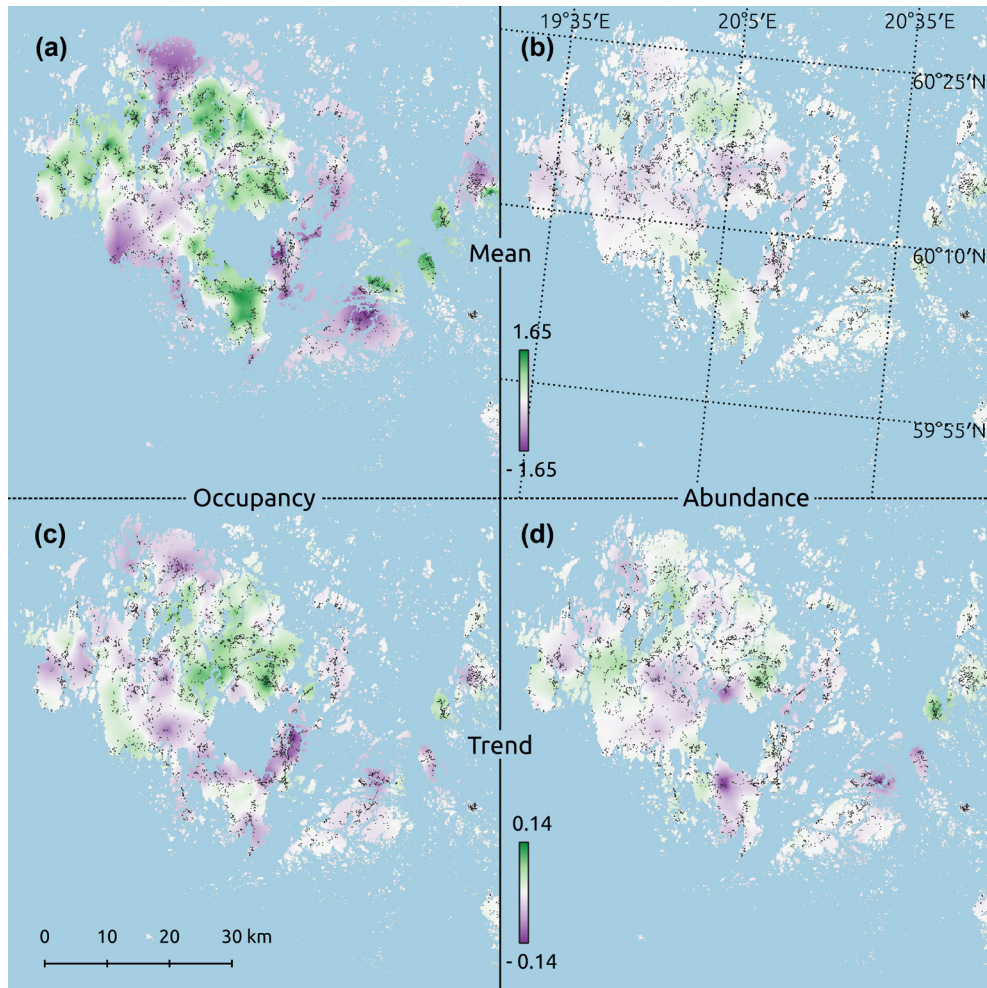
Figure 3. Summary of the spatio-temporal random fields for occupancy (a, c) and abundance (b, d). Panels (a), (b) show the mean of the spatial field over the years, panels (c–d) show the mean of annual changes of the spatial field over the years. We standardized the occupancy and abundance spatio-temporal random fields to unit variance before calculating the means and trends. Thus, the differences between occupancy and abundance are due to differences in the structure of the spatial field and not their differing magnitudes. Coastline data National Land Survey of Finland's (NLS) topographic database (acquired 08/2017). The grid lines in panel (b) show the position of the Åland Islands in the World Geodetic Coordinate System (WGS-84).

In the case of abundance, the strong role of the random effects in within-patch variation suggests that the covariates used do not capture the full range of variation in local abundance.

That unexplained variation can come in many guises such as unmeasured variation in landscape structure and habitat, or unaccounted interactions of the covariates with population dynamic processes such as when landscape structure affects dispersal (Conradt et al. 2000, Luoto et al. 2006, Stasek et al. 2008, DiLeo et al. 2018). Weather is another important missing factor, which affects butterfly population dynamics during most life stages (Heinrich 1993, Hellmann et al. 2004, Kuussaari et al. 2004) – also in the Åland Islands (Nieminen et al. 2004, Kahilainen et al. 2018). Interestingly, the spatio-temporal random field in the occupancy model is rather stable across the years, suggesting that the effects of climatic variation differ between occupancy and abundance. The effects on occupancy are limited to large-scale annual

differences with limited spatial variation, while for abundance the effects are more variable and also more local due to, for example, behavioural interactions with the local topography and microclimate (Eilers et al. 2013, Lawson et al. 2014). Any effects on abundance will also have indirect impacts on occupancy: for example, if summer drought reduces the quantity and quality of host plant available for larvae, the consequent reduction in abundance would also increase extinction risk in the next year (Salgado and Saastamoinen 2019).

**Within-population occupancy–abundance relationships**

In systems with sufficient turnover, such as classic metapopulations, occupancy–abundance relationships can be observed already at the level of local populations. In our system the qualitatively similar responses of occupancy and abundance
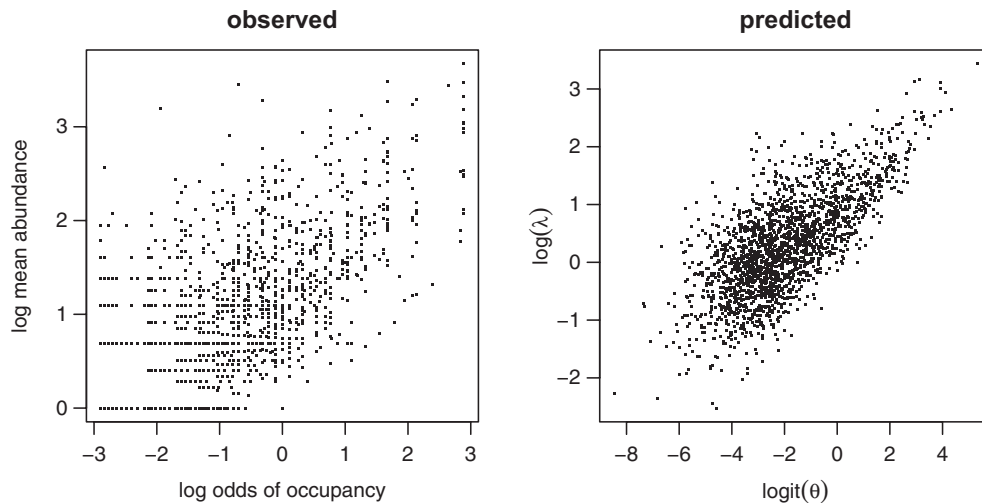
Figure 4. Observed and predicted within-patch occupancy–abundance relationships. The observed values are calculated for all patches occupied at least once, and their mean abundance is calculated only over the occupied years. The correlation between observed log odds and log abundance is 0.69 (n = 2022). The predicted values are patch means of the linear predictors calculated over all the observations (occupancy) or all occupancy observations (abundance). The correlation between the posterior means of the predicted log odds and predicted log abundance is 0.71 (n = 2027).

to environmental heterogeneity result in a noisy and non-linear, yet consistent, relationship between average occupancy and abundance within patches. In general, patches that have higher average occupancy have higher average abundance. This suggests that average occupancy can be a reasonable surrogate for mean abundance, which has also been shown in studies of other butterflies and species (Oliver et al. 2012, Gutiérrez et al. 2013). While it is harder to envision how negative within-population occupancy–abundance relationships would arise, lack of any clear relationship would suggest that the time series is too short compared to the lifespan of the study species or point to a drastic decline of habitat quality, such that overall average occupancy decreases and is independent of abundance before the decline.

A positive within-population occupancy–abundance relationship could follow from variation in growth rates or dispersal, mechanisms which are shown to generate positive within- and among-species occupancy–abundance relationships (Hanski and Gyllenberg 1993, Holt et al. 1997). Habitat patches with lower average growth rates have both smaller populations on average and go extinct more often; where immigration plays a strong role, low-quality or isolated habitat harbours occasional sink populations, while well connected habitat has inflated population sizes and high recolonization rates. In a wider context, the within-population relationship would then lead to a positive within-species occupancy–abundance relationship if mean growth rates or connectivity vary among different landscapes.

### Effect of abundance on occupancy

While occupancy is more predictable – in terms of variation explained by covariates – the large effect size of population size in the previous year suggests that the stochasticity

of abundance flows over into the occupancy model. While patch area and previous occupancy as surrogates for population size can suffice for predicting average occupancy – and thus average abundance – predictions of yearly changes in occupancy ignorant of population sizes would be biased towards the average case. Our result align with that of Oliver et al. (2012), who found that even when occupancy models predict mean population densities they do not capture long-term stability very well. Generally, this calls for caution when using occupancy as a surrogate for abundance in some contexts. Occupancy as means to discover environmental associations and to rank habitat in terms of quality is not very sensitive to assumptions about underlying relationship to abundance (Pearce and Ferrier 2001, Harrison et al. 2011). But trying to predict extinction dynamics of individual populations and estimate actual, not average, abundances using the same models could lead to biased estimates and higher uncertainty (Harrison et al. 2011, Bried and Pellet 2012). Thus, in the case of stochastic patch occupancy models, we should question these assumptions based on the intended use of the model as well as their biological plausibility (Hanski 1999, Gaggiotti and Hanski 2004, Ovaskainen and Hanski 2004a, b).

### Habitat quality and occupancy

Habitat quality, especially host plant abundance, has a strong role in determining occupancy: a quarter of all observations represent the lowest host plant abundance, but of occupied habitat only five percent fall into the lowest abundance class. This suggests a habitat quality threshold below which it is unlikely that patches become or stay occupied. Some of the possible explanations for such a threshold are that, the potential habitat patches are of such low quality that the butterfly

9

can only rarely, if ever, complete the larval stages of its life-cycle with the resources available, or that dispersing females are much less likely to find or choose low quality habitat, due to lack of suitable resources or absence of conspecifics (Kuussaari et al. 1996). A similarly stark contrast was found in relationship between host plant abundance and the presence of the Clouded Apollo (Luoto et al. 2001) and Oberthür's grizzled skipper (Fourcade and Öckinger 2017). In terms of notable land use effects, roads and surrounding agricultural lands increased occupancy, possibly due to increased connectivity between patches (DiLeo et al. 2018). In contrast, surrounding agriculture decreased abundance, possibly due to higher emigration rates (Kuussaari et al. 1996). These results reaffirm the need to account for population dynamics in attempting to study habitat quality and it's interaction with landscape context, both in terms of understanding the direct relationship between population density and habitat quality, as well as population density and emigration and immigration rates (Van Horne 1983).

### Analysis of variance

Our results underline that effect sizes and analysis of variance are not substitutes for each other, but complementary tools in observational studies, and must be interpreted in light of the ecology of the study system and the distribution and spatial structure of the covariates. For example, despite having a large effect size, having both host species present contributes little to overall variation, as within our study system in the Åland Islands one of the host plants, *Veronica spicata*, is not as widely distributed as the other, *Plantago lanceolata*; both are simultaneously present in only 16 percent of observations. Similarly, land use covariates, which show strong effects on occupancy, but due to a rather homogeneous landscape in areas where suitable habitat is found, explain very little variation in population dynamics. This would explain why Moilanen and Hanski (1998) did not find any model improvement from adding landscape composition to their stochastic patch occupancy model of the same system.

### Conclusions

Separating patterns of occupancy and abundance is the first step in determining the relationship between occupancy and abundance and how it responds to environmental heterogeneity. Given sufficiently reliable data, in terms of detection rate, the hurdle model provides a good basis for a statistical approach to this problem (Welsh et al. 1996). We emphasize the use of the hurdle model as a tool for meaningfully separating drivers of occupancy and abundance patterns, instead of concentrating on the so called zero inflation of the distribution of count data. We also advocate the use of spatio-temporal random effects to partially compensate for the effects of suitable, but unoccupied, habitat on covariate coefficient estimates, in the presence of dispersal and potential unmeasured covariates.

In our system, conditional on occupancy, unexplained variation in abundance has no strong spatial structure, while the unexplained spatio-temporal variance in occupancy is rather stable when excluding noise due to variation in abundance. This suggests that in some cases spatial structure estimated in abundance-only species distribution models would lead to a compromise mixing two different patterns that arise from partially different processes.

A positive relationship between average occupancy and abundance within local populations can arise from the relationship between local abundance and extinction rate, or immigration and the extinction and colonization rates. The relationship we observed is noisy and non-linear, and suggests that in spite of the consistency of the relationships, occupancy is not always a good surrogate for abundance in a dynamic modelling context

### Data availability statement

Data and code available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.ksn02v707> (Schulz et al. 2019).

### References

Araújo, M. B. et al. 2005. Validation of species–climate impact models under climate change. – Global Change Biol. 11: 1504–1513.

Aue, B. et al. 2012. Distance weighting avoids erroneous scale effects in species-habitat models. – Methods Ecol. Evol. 3: 102–111.

Bried, J. T. and Pellet, J. 2012. Optimal design of butterfly occupancy surveys and testing if occupancy converts to abundance for sparse populations. – J. Insect Conserv. 16: 489–499.

Buckley, L. B. et al. 2010. Can mechanism inform species' distribution models? – Ecol. Lett. 13: 1041–1054.

Cameron, S. A. et al. 2011. Patterns of widespread decline in North American bumble bees. – Proc. Natl Acad. Sci. USA 102: 662–667.

Chandler, R. and Hepinstall-Cymerman, J. 2016. Estimating the spatial scales of landscape effects on abundance. – Landscape Ecol. 28: 1383–1394.

Conrad, K. F. et al. 2006. Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis. – Biol. Conserv. 132: 279–291.

Conradt, L. et al. 2000. Non-random dispersal in the butterfly *Maniola jurtina*: implications for metapopulation models. – Proc. R. Soc. B 267: 1505–1510.

Cowley, M. J. R. et al. 2001. Density–distribution relationships in British butterflies. I. The effect of mobility and spatial scale. – J. Anim. Ecol. 70: 410–425.

Dallas, T. A. and Hastings, A. 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. – Global Ecol. Biogeogr. 27: 1448–1456.

DiLeo, M. F. et al. 2018. Landscape permeability and individual variation in a dispersal-linked gene jointly determine genetic structure in the Glanville fritillary butterfly. – Evol. Lett. 2: 544–556.

Duff, T. J. et al. 2012. Patterns of plant abundances in natural systems: is there value in modelling both species abundance and distribution? – Aust. J. Bot. 59: 719–733.

Eilers, S. et al. 2013. Micro-climate determines oviposition site selection and abundance in the butterfly *Pyrgus armoricanus* at its northern range margin. – Ecol. Entomol. 38: 183–192.

Etienne, R. S. et al. 2004. Application of stochastic patch occupancy models to real metapopulations. – In: Hanski, I. and Gaggiotti, O. E. (eds), Ecology, genetics and evolution of metapopulations. Academic Press, pp. 105–132.

Fourcade, Y. and Öckinger, E. 2017. Host plant density and patch isolation drive occupancy and abundance at a butterfly's northern range margin. – Ecol. Evol. 7: 331–345.

Freckleton, R. P. et al. 2005. Large-scale population dynamics, abundance–occupancy relationships and the scaling from local to regional population size. – J. Anim. Ecol. 74: 353–364.

Freckleton, R. P. et al. 2006. Distributions of habitat suitability and the abundance–occupancy relationship. – Am. Nat. 167: 260–275.

Gaggiotti, O. E. and Hanski, I. 2004. Mechanisms of population extinction. – In: Hanski, I. and Gaggiotti, O. E. (eds), Ecology, genetics and evolution of metapopulations. Academic Press, pp. 337–366.

Gaston, K. J. et al. 2000. Abundance–occupancy relationships. – J. Appl. Ecol. 37: 39–59.

Gelman, A. et al. 2014. Bayesian data analysis, 3rd ed. – CRC Press.

Georgina, P. et al. 2015. Individualistic sensitivities and exposure to climate change explain variation in species' distribution and abundance changes. – Sci. Adv. 1: e1400220.

Georgina, P. et al. 2016. Retraction of the research article: 'individualistic sensitivities and exposure to climate change explain variation in species' distribution and abundance changes'. – Sci. Adv. 2: e1600819.

Gutiérrez, D. et al. 2013. Models of presence–absence estimate abundance as well as (or even better than) models of abundance: the case of the butterfly *Parnassius apollo*. – Landscape Ecol. 28: 401–413.

Hallmann, C. A. et al. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. – PLoS One 12: e0185809.

Hanski, I. 1991. Single-species metapopulation dynamics: concepts, models and observations. – Biol. J. Linn. Soc. 42: 17–38.

Hanski, I. 1994. A practical model of metapopulation dynamics. – J. Anim. Ecol. 63: 151–162.

Hanski, I. 1999. Metapopulation ecology. – Oxford Univ. Press.

Hanski, I. and Gyllenberg, M. 1993. Two general metapopulation models and the core-satellite species hypothesis. – Am. Nat. 142: 17–41.

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

Hanski, I. et al. 1994. Metapopulation structure and migration in the butterfly *Melitaea cinxia*. – Ecology 75: 747–762.

Hanski, I. et al. 1995. Metapopulation persistence of an endangered butterfly in a fragmented landscape. – Oikos 72: 21–28.

Hanski, I. et al. 2017. Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. – Nat. Commun. 8: 14504.

Harrison, P. J. et al. 2011. Bayesian state-space modeling of metapopulation dynamics in the Glanville fritillary butterfly. – Ecol. Monogr. 81: 581–598.

Heinrich, B. 1993. Hot-blooded insects: strategies and mechanisms of thermoregulation in insects. – Springer.

Hellmann, J. J. et al. 2004. Structure and dynamics of *Euphydryas editha* populations. – In: Erlich, P. R. and Hanski, I. (eds), On the wings of checkerspots: a model system for population biology. Oxford Univ. Press, pp. 34–62.

Holt, A. R. et al. 2002. Occupancy–abundance relationships and spatial distribution: a review. – Basic Appl. Ecol. 3: 1–13.

Holt, R. D. et al. 1997. On the relationship between range size and local abundance: back to basics. – Oikos 78: 183–190.

Hyvärinen, E. et al. 2019. The 2019 Red List of Finnish species. – Finnish Environment Inst.

Kahilainen, A. et al. 2018. Metapopulation dynamics in a changing climate: increasing spatial synchrony in weather conditions drives metapopulation synchrony of a butterfly inhabiting a fragmented landscape. – Global Change Biol. 24: 4316–4329.

Kallasvuo, M. et al. 2017. Modeling the spatial distribution of larval fish abundance provides essential information for management. – Can. J. Fish. Aquat. Sci. 74: 636–649.

Keeling, M. J. 2002. Using individual-based simulations to test the Levins metapopulation paradigm. – J. Anim. Ecol. 71: 270–279.

Keith, D. A. et al. 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. – Biol. Lett. 4: 560–563.

Koh, I. et al. 2016. Modeling the status, trends and impacts of wild bee abundance in the United States. – Proc. Natl Acad. Sci. USA 113: 140–145.

Komonen, A. et al. 2019. Alarmist by bad design: strongly popularized unsubstantiated claims undermine credibility of conservation science. – Rethink. Ecol. 4: 17–19.

Kuussaari, M. and Singer, M. C. 2017. Group size and egg and larval survival in the social butterfly *Melitaea cinxia*. – Ann. Zool. Fenn. 54: 213–223.

Kuussaari, M. et al. 1996. An experimental study of migration in the Glanville fritillary butterfly *Melitaea cinxia*. – J. Anim. Ecol. 65: 791–801.

Kuussaari, M. et al. 2004. Larval biology of checkerspots. – In: Erlich, P. R. and Hanski, I. (eds), On the wings of checkerspots: a model system for population biology. Oxford Univ. Press, pp. 138–160.

Lawson, C. R. et al. 2014. Topographic microclimates drive microhabitat associations at the range margin of a butterfly. – Ecography 37: 732–740.

Leroux, S. J. et al. 2013. Mechanistic models for the spatial spread of species under climate change – Ecol. Appl. 23: 815–828.

Lewthwaite, J. M. M. et al. 2018. Canadian butterfly climate debt is significant and correlated with range size. – Ecography 41: 2005–2015.

Lindén, A. and Mäntyniemi, S. 2011. Using the negative binomial distribution to model overdispersion in ecological count data. – Ecology 92: 1414–1421.

Lindgren, F. and Rue, H. 2015. Bayesian spatial modelling with R-INLA. – J. Stat. Softw. 63: 1–25.

Lindgren, F. et al. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. – J. R. Stat. Soc. B 73: 423–498.

Luoto, M. et al. 2001. Determinants of distribution and abundance in the clouded Apollo butterfly: a landscape ecological approach. – Ecography 24: 601–617.

Luoto, M. et al. 2006. Determinants of the biogeographical distribution of butterflies in boreal regions. – J. Biogeogr. 33: 1764–1778.

Mäkisara, K. et al. 2016. The multi-source national forest inventory of Finland – methods and results 2013. – Natural Resources Inst. Finland.

Martin, A. E. and Fahrig, L. 2012. Measuring and selecting scales of effect for landscape predictors in species–habitat models. – Ecol. Appl. 22: 2277–2292.

Miguet, P. et al. 2017. How to quantify a distance-dependent landscape effect on a biological response. – Methods Ecol. Evol. 8: 1717–1724.

Moilanen, A. and Hanski, I. 1998. Metapopulation dynamics: effects of habitat quality and landscape structure. – Ecology 79: 2503–2515.

Nieminen, M. et al. 2004. Structure and dynamics of *Melitaea cinxia* populations. – In: Erlich, P. R. and Hanski, I. (eds), On the wings of checkerspots: a model system for population biology. Oxford Univ. Press, pp. 63–91.

Ojanen, S. P. et al. 2013. Long-term metapopulation study of the Glanville fritillary butterfly (*Melitaea cinxia*): survey methods, data management and long-term population trends. – Ecol. Evol. 3: 3713–3737.

Oliver, T. H. et al. 2012. Population density but not stability can be predicted from species distribution models. – J. Appl. Ecol. 49: 581–590.

Ovaskainen, O. and Hanski, I. 2004a. From individual behavior to metapopulation dynamics: unifying the patchy population and classic metapopulation models. – Am. Nat. 164: 364–377.

Ovaskainen, O. and Hanski, I. 2004b. Metapopulation dynamics in highly fragmented landscapes. – In: Hanski, I. and Gaggiotti, O. E. (eds), Ecology, genetics and evolution of metapopulations. Academic Press, pp. 73–103.

Ovaskainen, O. and Saastamoinen, M. 2018. Frontiers in metapopulation biology: the legacy of Ilkka Hanski. – Annu. Rev. Ecol. Evol. Syst. 49: 231–252.

Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – Ecol. Lett. 20: 561–576.

Pearce, J. and Ferrier, S. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. – Biol. Conserv. 98: 33–43.

Rue, H. et al. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. – J. R. Stat. Soc. B 71: 319–392.

Salgado, A. L. and Saastamoinen, M. 2019. Developmental stage-dependent response and preference for host plant quality in an insect herbivore. – Anim. Behav. 150: 27–38.

Sánchez-Bayo, F. and Wyckhuys, K. A. 2019. Worldwide decline of the entomofauna: a review of its drivers. – Biol. Conserv. 232: 8–27.

Schulz, T. et al. 2019. Data from: Long-term demographic surveys reveal a consistent relationship between average occupancy and abundance within local populations of a butterfly metapopulation. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.ksn02v707>.

Shortall, C. R. et al. 2009. Long-term changes in the abundance of flying insects. – Insect Conserv. Divers. 2: 251–260.

Sileshi, G. 2007. A method for estimating insect abundance and patch occupancy with potential applications in large-scale monitoring programmes. – Afr. Entomol. 15: 89–101.

Sileshi, G. et al. 2009. Traditional occupancy–abundance models are inadequate for zero-inflated ecological count data. – Ecol. Model. 220: 1764–1775.

Stasek, D. J. et al. 2008. Butterfly abundance and movements among prairie patches: the roles of habitat quality, edge and forest matrix permeability. – Environ. Entomol. 37: 897–906.

Thomas, C. D. et al. 2019. 'Insectageddon': a call for more robust data and rigorous analyses. – Global Change Biol. 25: 1891–1892.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. – J. Wildl. Manage. 47: 893–901.

Welsh, A. et al. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. – Ecol. Model. 88: 297–308.

Yuan, Y. et al. 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. – Ann. Appl. Stat. 11: 2270–2297.