Doors and corners of variance partitioning in statistical ecology

Running head: Variance partitioning in ecology

Authors: Torsti Schulz[1]

Marjo Saastamoinen[1,2]

Jarno Vanhatalo[1,3]

Affiliations: [1] Organismal and Evolutionary Biology Research Programme, University of Helsinki

[2] Helsinki Institute of Life Science HiLIFE, University of Helsinki

[3] Department of Mathematics and Statistics, University of Helsinki

Corresponding author: Torsti Schulz

Abstract

Variance partitioning is a common tool for statistical analysis and interpretation in both observational and experimental studies in ecology. Its popularity has led to a proliferation of methods with sometimes confusing or contradicting interpretations. Here, we present variance partitioning as a general tool in a model based Bayesian framework for summarizing and interpreting regression-like models. To demonstrate our approach we present a case study comprising of a simple occupancy model for a metapopulation of the Glanville fritillary butterfly. We pay special attention to the thorny issue of correlated covariates and random effects, and highlight uncertainty in variance partitioning. We recommend several alternative measures of variance, which jointly can be used to better interpret variance partitions. Additionally, we extend the general approach to encompass partitioning of variance within and between groups of observations, an approach very similar to analysis of variance. While noting that many troublesome issues relating to variance partitioning, such as uncertainty quantification, have been neglected in the literature, we likewise feel that the rather general applicability of the methods as an extension of statistical model-based analyses has not been fully utilized by the ecological research community either.

Keywords: variance partitioning, model-based inference, uncertainty quantification

1 Introduction

In ecology, we are often interested in determining the processes and conditions that modify or maintain the distribution of species and the composition of ecological communities (Hubbell 2001, Ives *et al*. 2003, Vellend 2010, Ovaskainen *et al*. 2016). Key questions are what abiotic and biotic variables affect individual species and communities, what have been the relative roles of these variables in the past, and what will they be in the future (e.g. MacKenzie *et al*. 2007, Peres-Neto *et al*. 2006, Estay *et al*. 2014, Vuorinen *et al*. 2015, Kahilainen *et al*. 2018). These questions are intertwined but still largely separate. For example, an environmental variable such as temperature can affect species survival and reproduction rate in general but, if the study population experiences only moderate temperature variability, the contribution of temperature to variation in population size might be negligible in a particular study site (see *e.g.* Tomkiewicz *et al*. 1998, Veneranta *et al*. 2011). It is partially this distinction between potential variation and realized variation that has motivated our work and probably has contributed to the large array of methods of variance partitioning.

In experimental research, the relative importance of alternative variables in explaining variation in a response variable is traditionally studied using the analysis of variance (ANOVA) approach, which provides a direct method for separating background variation from treatment effects under controlled conditions (Casella and Berger 2002, Gelman 2005). In theoretical studies, the relative importance of alternative driving processes can be examined by mechanistic (first principles) models either by analytically studying their behaviour or by simulation (Ovaskainen *et al*. 2016). However, mechanistic models and large ecological experiments that would be necessary to directly study ecological processes at large scales and over longer time spans are feasible only in limited cases (*cf*. Cressie and Wikle 2011, Péron and Koons 2012, McCallum 2016). Consequently, empirical studies usually rely on observational data, often combined with model based statistical inference (Mac Nally 1996, Ives *et al*. 2003). For example, species distribution

models and a variety of regression-based approaches are used to associate effects of covariates and structured random effects to the presence and abundance of species or the composition communities.

In these circumstances, where direct interpretation of effect sizes is hampered by lack of experimental control and correlations among model predictors, statistical variance partitioning methods enter the stage as an aid for interpreting the results of these regression-like models (Pedhazur 1997, pp. 241–282, Legendre and Legendre 2012). Variance partitioning is used towards various ends, such as understanding how variation is distributed at different hierarchical levels in nested experiments or multi-scale studies (Searle *et al*. 1992, Cushman and McGarigal 2002, Goldstein *et al*. 2002, McMahon and Diez 2007) and determining what is the relative importance of a covariate in terms of model performance or contribution to the process studied (Chevan and Sutherland 1991, Bring 1995, Mac Nally 1996). In other cases, the interest lies simply in separating the unique and shared contributions of covariates (Ray-Mukherjee *et al*. 2014). We also find methods such as ordination (Borcard *et al*. 1992), standardized regression coefficients (Bring 1994, 1995), variance partitioning coefficients (Goldstein *et al*. 2002), and factor analysis (Riitters *et al*. 1995) being used for variance partitioning.

The abundance of applications and methods go hand in hand with the ad-hoc nature of many approaches to variance partitioning and determination of relative importance of covariates. Often, the use of variance partitioning arises from a specific need, or for a specific kind of model. In that applied context the potential for rather wide applicability of the general approach does not become evident. These applications may also take shortcuts that do not respect the model structure or do not directly answer the question they are expected to answer (Pedhazur 1997, pp. 241–282). Furthermore, clear recommendations for how to interpret results and what should be reported when utilizing variance partitioning are lacking. To give an example, only a few studies report uncertainties in variance partition estimates or indicate the strength of

correlations between variance components (but see Peres-Neto *et al*. 2006, Ray-Mukherjee *et al*. 2014, Yuan *et al*. 2017, Schulz *et al*. 2020).

As a further challenge, statistical approaches rarely fulfill the dream of causal understanding (Wright 1921, Burks 1926, Levins and Lewontin 1982, Cooper 1998, Cressie and Wikle 2011), and meaningful interpretation of estimated effect sizes of covariates and variance parameters of so-called random effects is challenging when modelling observational data. While at best the effect sizes reflect the actual importance of the covariates on the species or community studied, even then they might not reflect how important these covariates are for predicting variability in population size or community composition over space or time. The importance of the covariates in any specific context depends not only on the existence of some underlying, and usually unconfirmed, causal relationship but also on the variability in the covariates themselves.

To give a concrete example, the Baltic Sea is a brackish water basin hosting partially endemic fish species, such as Baltic herring, sprat, pikeperch, and turbot, which are all sensitive to water salinity. Even though the salinity gradient is large throughout the Baltic Sea, salinity predicts only very little of the interannual variation in the species' reproduction rates as there is generally little interannual variation in water salinity in regions where these fish species occur (e.g. Lappalainen and Lehtonen 1995, Casini *et al*. 2006, Nissling *et al*. 2006). It is also often the case that any of the survey data at hand do not represent the full extent of variation in the covariates in the areas where the species occur (Vernier *et al*. 2008, *cf*. Foster *et al*. 2021). In either case, if data contain little or no variation with respect to a covariate, then there is no statistical way to assess its effect and thus no variation in the outcome can be attributed to that covariate, whether it is important or not (Pratt 1987). In extreme cases the situation is akin to the lack of genetic variability in an allele fixed either due to its high fitness advantage or simply due to genetic drift (*cf*. Whitlock and Gomulkiewicz 2005).

Keeping the preceding admonitions in mind, we follow Gelman (2005, Gelman *et al*. 2014, 381–400, Gelman *et al*. 2019) to present a model based approach to variance partitioning. Our goal is to quantify the contributions of individual model components to variation in the quantity of interest, such as species abundance or community composition. A similar Bayesian approach has been utilized by *e.g*. Yuan et al. (2017), who used it as a model checking tool for a spatio-temporal species distribution model, by Schulz *et al*. (2020) where we used it to explore occupancy–abundance relationships in a butterfly metapopulation, and in the joint species distribution modelling software HMSC (Ovaskainen and Abrego 2020, Tikhonov *et al*. 2020). Instead of focusing only on a specific application, we present variance partitioning as a more general tool for the analysis and interpretation of statistical models. We also demonstrate how we can avoid some of the pitfalls of conducting and interpreting variance partitioning studies present in earlier work as well as show its applicability to predictive scenarios. We specifically address topics such as correlations between covariates, confounding by random effects and correlated parameters, which have received only little attention in the contemporary literature (but see Wakefield 2003, 2004, Yuan *et al*. 2017). Moreover, we propose variance partitioning for grouping of covariates as well as conditioning on observations and present some common as well as new visualization techniques for variance partitions. We derive and discuss the methods in a generalized linear model setting but variance partitioning is, at least partially, applicable to a wider class of statistical models, namely any hierarchical additive models (Gelman 2005), including traditional GLMs and GAMs as well as their extensions with structured and unstructured random effects (Hodges *et al*. 2007).

The rest of the paper is organized as follows. We first introduce the basic concepts behind variance partitioning in Section 2.1, after which we introduce the theory for its extensions in sections 2.2–2.7. In Section 3, we demonstrate the different variance partitioning techniques with a case study using a simple species distribution modeling (SDM) -like regression model. We will apply it to model presence–absence of an insect herbivore in a metapopulation context. After

Section 2.1, practically oriented readers can jump directly to the case study and refer to the theory after reading the motivating examples. For researchers interested in applying these methods we also provide worked examples including all necessary code in the Programming supplement.

2 Material and methods

2.1 Introduction to variance partitioning

To better understand how variance partitioning works and is easiest to interpret, we first look at an idealized scenario from population ecology, where a closed population reproduces according to a simple time-varying Ricker population model (Ricker 1954). The Ricker model for population growth at time t with population size N_t can be written as (Ruppert and Carroll 1985, Collie *et al*. 2012, Weigel *et al*. 2021)

$$\log \frac{N_{t+1}}{N_t} = y_t = \alpha + \beta_1 x_t + \beta_2 N_t + \epsilon_t \tag{1}$$

where $\alpha + \beta_1 x_t$ stands for the intrinsic population growth as a function of environmental covariates x_t, β_2 (< 0) is a parameter governing density-dependent survival and ϵ_t models intrinsic stochasticity in the population dynamics. Thus, we can describe the log growth rate y_t of this idealized population with a linear model, whose parameters have an intuitive ecological interpretation. Hence, we would hope calculating and interpreting the variance partitioning were straightforward as well. However, this is not the case in practice.

Let us first assume that we know the parameters of the model, $\theta = [\alpha, \beta_1, \beta_2]$, and we can measure the population size and environmental covariates accurately for years $t = 1, \dots, n$. The standard approach to analyze what proportion of the variation in log-population growth rate arises from the three different processes described in our model would be to calculate

$$V_x(\theta) = \frac{\mathrm{Var}[\beta_1 x]}{\mathrm{Var}[y]}, V_N(\theta) = \frac{\mathrm{Var}[\beta_2 N]}{\mathrm{Var}[y]}, V_\epsilon(\theta) = \frac{\mathrm{Var}[\epsilon]}{\mathrm{Var}[y]} \quad (2)$$

where the variances are sample variances over the n observations collected in vectors, such as $y = [y_1, \dots, y_n]^{\mathrm{T}}$.

In order for these terms to summarize the *proportions* of variation they should sum up to one. However they sum up to one, only in the special case that the three components of the Ricker model are mutually independent so that $\mathrm{Var}[y] = \mathrm{Var}[\beta_1 x] + \mathrm{Var}[\beta_2 N] + \mathrm{Var}[\epsilon]$. This is an unrealistic assumption in most cases since, for example, temporal correlation in environmental covariates ($\mathrm{Cov}[x_t, x_{t-1}] \neq 0$) induces correlation between the intrinsic growth rate and the density dependent population dynamics (Mathematical supplement 1, see also Ives *et al*. 2003). If the correlation between $\beta_1 x_t$ and N_t is positive then $\mathrm{Var}[\beta_1 x] + \mathrm{Var}[\beta_2 N] + \mathrm{Var}[\epsilon] < \mathrm{Var}[y]$ and if they are negatively correlated, then the sum of variance terms is larger than the variance of y.

In ecological applications, it is common to have temporally or spatially correlated environmental covariates and we refer to this phenomenon as the *correlated covariates challenge* (see also Cox and Snell 1974, Pedhazur 1997, pp. 241–282, Graham 2003, Dormann *et al*. 2013, Morrisey and Ruxton 2018). The example of fisheries in the Baltic Sea above serves to illustrate the problem: salinity, water temperature, and depth can all play a role in determining spawning areas for many of the fish species, but they are hard to disentangle as they are also all highly correlated with each other (Tomkiewicz *et al*. 1998, Janssen *et al*. 1999, MacKenzie *et al*. 2007, Veneranta *et al*. 2016).

A related challenge that has received less attention in the ecological literature is the challenge posed by correlation between the random effects and the covariates (Wakefield 2004, Yuan *et al*. 2017). The random effects ϵ_t are typically defined to be *a priori* independent from the other model components. However, most often our model does not contain all relevant covariates, but

we may have missed a covariate z_t that is correlated with x_t. When estimating the random effects from data, the effect of this missing covariate will be captured by ϵ_t at least partially, leading to correlation between $\beta_1 x_t$ and ϵ_t. Correlation between $\beta_1 x_t$ and ϵ_t can also arise for other reasons such as when a model includes both a spatial random effect and covariates with spatial structure (Teng *et al*. 2018). We call this phenomenon the *confounded random effects challenge*.

One approach to surpass the correlated covariates and confounded random effects challenges is to calculate the variance partitioning as

$$P_x(\theta) = \frac{\mathrm{Var}[\beta_1 x]}{\mathrm{Var}[\beta_1 x] + \mathrm{Var}[\beta_2 N] + \mathrm{Var}[\epsilon]} \tag{3}$$

which ensures that the different variance components sum up to one by ignoring the covariances in the denominator (Ovaskainen and Abrego 2020, pp. 69–70). We call this the *diagonal variance partition* as when looking at the covariance matrix of the components, this measure only accounts for the diagonal variation. The diagonal variance partition is only valid when the correlations are negligible or of no consequence to the interpretation of the results, so it is at most a good approximation for a 'true' variance contribution of the covariate. Hence, this diagonal approximation is not a general solution to the problem.

Up to this point, we have assumed that the model parameters are known. Naturally, this is usually not the case, since if it was, we would not need to do statistical inference in the first place. In this work, we follow the Bayesian approach and calculate the full posterior distribution for model parameters $p(\theta|D)$ where $D = \{y_t, x_t, N_t\}_{t=1}^{n}$ collects the data and $\theta = [\alpha, \beta_1, \beta_2]$ the parameters, as above. The posterior distribution for the parameters induces a posterior distribution for the variance partitioning statistics, so that instead of a single estimate for $V(\theta)$, the variance partitioning, we have a random variable (or rather random vector) with distribution

$V \sim p(V|D)$ (*cf.* Gelman *et al*. 2019). From a Bayesian point of view, it is natural to study this posterior distribution which we can use as a measure of the uncertainty in our variance partitions. The uncertainty in variance partitioning is not, however, typically reported but what is traditionally done (e.g. Cushman and McGarigal 2002, Legendre and Legendre 2012) is that people summarize the model parameters with a point estimate $\hat{\theta}$, such as the posterior mean, or more commonly the least squares or maximum likelihood estimate, and report $V(\hat{\theta})$ or $P(\hat{\theta})$ or some similar measure (but see *e.g*. Goldstein *et al*. 2002, Yuan *et al*. 2017). These estimates suffer from the correlated covariates and confounded random effects challenges as discussed above.

However, there is even more to the story. In addition to neglecting the uncertainty in variance partitioning, there is no guarantee that $V(\hat{\theta})$ and $P(\hat{\theta})$ are actually good summaries for $p(V|D)$ and $p(P|D)$. Even if the parameters were uncorrelated *a posteriori*, directly substituting point estimates into the formulas will be erroneous. Take for example the variance partition of the covariate effect $V_x(\theta) \propto \mathrm{Var}[\beta_1 x]$ (the $\propto$ denotes "is proportional to"). The plug-in estimate $V_x(\hat{\theta}) = \mathrm{Var}[\hat{\beta}_1 x] = \hat{\beta}_1^2 \mathrm{Var}[x]$ where $\hat{\beta}_1 = E[\beta_1|D]$ (the posterior mean) would not lead to the posterior mean $\mathrm{E}[V_x|D]$ since $\mathrm{E}[\beta_1^2|D] > \hat{\beta}_1^2$ unless the variance of β_1 is zero (Jensen's inequality). In the case of P_x the difference between the plug-in estimate $P_x(\hat{\theta})$ and $\mathrm{E}[P_x|D]$ is typically even larger, since also its denominator is a function of parameters (see eq. (3)). This point is especially salient in the Bayesian context, but can apply equally to some frequentist measures of uncertainty about the point estimate of variance partitioning.

The correct approach to calculate a point summary for the mean and spread of a variance partition is using the distributions $p(V|D)$ and $p(P|D)$, most generally using the expectation and variance of their distribution

$$\hat{V} = \mathrm{E}[V(\theta)], \qquad \mathrm{Var}[V(\theta)] = \mathrm{E}[(V(\theta) - \mathrm{E}[V(\theta)])^2] \tag{4}$$

$$\hat{P} = \mathrm{E}[P(\theta)], \qquad \mathrm{Var}[P(\theta)] = \mathrm{E}[(P(\theta) - \mathrm{E}[P(\theta)])^2]$$

These summaries can easily be approximated using samples from the posterior distribution of $p(\theta|D)$ as it is sufficient to know the distribution of the parameters θ to calculate the expectation of functions of these parameters (Casella and Berger 2002, p. 55; see also Section 3, Mathematical supplement 2.7, and the Programming supplement). This approach avoids the bias that substituting point estimates into the formulas directly introduces and is used for example in HMSC (Table 1). However, if we then only communicate the posterior means $\hat{V}$ or $\hat{P}$, we lose knowledge of how the uncertainty in model parameters propagates to the variance partitions. The least should be to also report the posterior variance of the estimates.

In the Bayesian context, we must also consider what happens when the model parameters are mutually correlated in their posterior distribution, that is, for example, $\mathrm{Cor}[\beta_1, \beta_2|D] \neq 0$. The posterior correlations between parameters translate into posterior correlations between the variance partition measures V_x and V_N or equivalently P_x and P_N. In that case, even the marginal distributions of the variance partitions are not necessarily good summaries of their joint posterior distribution (see Sections 2.6 and 3.6).

2.2 Variance partitioning of the linear predictor

Moving on to a general regression context, we discuss variance partitioning in the context of the *linear predictor*. In a generalized linear model, the linear predictor is $\eta = g\big(E(y)\big)$, a transformation of the expected value of outcomes y to the linear space using a link function g. Linear predictor is a length n vector, that is formed as the sum of d vectors of *linear terms*, that is $\eta(x) = \sum_{j=1}^{d} x_j \beta_j$ where , $x = \{x_j\}_{j=1}^{d}$ and x_j is a length n vector of covariates. But as far as variance partitioning is concerned the linear terms could equally be additive or random effects. The variance over the n observations of the linear predictor in (generalized) linear models is

$\mathrm{Var}[\eta] = \mathrm{Var}\left[\sum_{j=1}^{d} x_j\beta_j\right]$. Hence its decomposition follows from the properties of variances of linear combinations (Mulaik 2010, pp. 73–76, 83–84) and can be written as a sum of the covariances of the linear terms:

$$\mathrm{Var}[\eta] = \sum_{j=1}^{d}\sum_{j'=1}^{d} \mathrm{Cov}[x_j\beta_j, x_{j'}\beta_{j'}] \tag{5}$$

Here our variance partition measures are $V_{x_j}(\theta) = \frac{\mathrm{Var}[x_j\beta_j]}{\mathrm{Var}[\eta]}$ and $P_{x_j}(\theta) = \frac{\mathrm{Var}[x_j\beta_j]}{\sum_{j'=1}^{d}\mathrm{Var}\left[x_{j'}\beta_{j'}\right]}$ for the jth covariate x_j. To see how this generalizes to any generalized additive model with random effects, remember that if we write the linear terms in the form $f_j(x_j)$ then $\eta(x) = \sum_{j=1}^{d} f_j(x_j)$, and we have an unweighted linear combination of the additive functions or random effects and hence can still apply all the methods presented in this paper (i.e. $V_{x_j} = \frac{\mathrm{Var}[f_j(x_j)]}{\mathrm{Var}[\eta]}$). We could also calculate the variance partitioning for non-linear joint effects of two or several covariates, such as $f(x_j, x_{j'})$. This is used, for example, when calculating the variance contribution of continuous spatial random effects.

The approach is not limited to partitioning the variation in the linear predictor, but can also be used for the more traditional partitioning of explained variation, R^2, when working with a linear model or a model with linearizable residuals. If we partition variance while accounting for residual variance, it will give us a measure that is equal to the traditional R^2 when defined as $\frac{\mathrm{Var}[\eta]}{\mathrm{Var}[\eta]+\mathrm{Var}[\epsilon]}$. The only change for the variance partitioning metrics $V_x(\theta)$ and $P_x(\theta)$ is the inclusion of the residual variance in the denominator so that $V_{x_j}(\theta) = \frac{\mathrm{Var}[x_j\beta_j]}{\mathrm{Var}[\eta+\epsilon]}$ and $P_{x_j}(\theta) = \frac{\mathrm{Var}[x_j\beta_j]}{\sum_{j'=1}^{d}\left(\mathrm{Var}\left[x_{j'}\beta_{j'}\right]\right)+\mathrm{Var}[\epsilon]}$. The residuals ϵ can be calculated either with respect to the observations y or its posterior predictive distribution $\tilde{y}$ as is done in the Bayesian R^2 approach (Gelman *et al.* 2019) but we do not consider this choice further here.

2.3 Grouping linear terms in variance partitioning

Often analyses include many environmental covariates, which can be divided into multiple groups, such as habitat quality or climatic variables. These *groups of linear terms* are sometimes called *predictor groups* (Ovaskainen and Abrego 2020, pp. 69–70) or simply *sets of variables* in some ecological literature (e.g. Legendre and Legendre 2012, *cf.* Wisler 1969, Beaton 1973, Cox and Snell 1974). We can calculate the variance partitioning at the level of these groups of linear terms by summing over the variance terms of the covariates included in that group (*cf.* Mulaik 2010, pp. 72–84). To present this explicitly, we define the following notation. If each covariate is associated with only one of m groups, the linear term group effects can be written formally as $b_l = \sum_{j \in B_l} x_j \beta_j$ where $l = 1, \dots, m$ and B_l is a set of covariate indexes that are associated with the l'th group. We can rewrite the linear predictor in terms of sums of the linear term group effects, $\eta = \sum_{l=1}^{m} b_l$. The variance of a group effect is $\mathrm{Var}[b_l] = \mathrm{Var}\left[\sum_{j \in B_l} x_j \beta_j\right] = \sum_{j \in B_l} \sum_{j' \in B_l} \mathrm{Cov}\left[x_j \beta_j, x_{j'} \beta_{j'}\right]$ and can be used to express the variance of the linear predictor $\mathrm{Var}[\eta] = \sum_{l=1}^{m} \sum_{l'=1}^{m} \mathrm{Cov}[b_l, b_{l'}]$. The linear predictor group variance partition is then

$$V_{b_l}(\theta) = \frac{\mathrm{Var}[b_l]}{\mathrm{Var}[\eta]} \text{ and } P_{b_l}(\theta) = \frac{\mathrm{Var}[b_l]}{\sum_{l'=1}^{m} \mathrm{Var}[b_{l'}]} \tag{6}$$

for some group b_l (Mathematical supplement 2.3).

In addition to simplifying the presentation of models with many covariates by summarizing them at a more abstract thematic level, it can also help interpreting models with correlated covariates when the correlations between covariates are greater within the linear term groups than between them: grouping related covariates incorporates the covariances between those covariates to the variance term of that group and thus reduces the covariance between the (grouped) variance components (Ovaskainen and Abrego 2020, pp. 69–70). This can also reduce the posterior correlation between the variance partitions (see Section 2.6) compared to a situation without

groups. However, this does not solve either of the problems entirely, since correlations can occur also between thematically unrelated covariates leading to covariances and posterior correlations between the groups.

It is to be noted, that the grouping of the covariates is not a property of the observations, or the current sample of data at hand, but rather a choice made by the analyst as will be illustrated in the case study (Section 3.1). It thus incorporates external knowledge into the interpretation of results. Also, depending on how the groups of linear terms are defined, a group might consist of a mixture of covariate and random effects (*cf.* Schulz et al. 2020).

2.4 Variance partitioning over populations and its use in prediction

The approach to variance partitioning presented thus far corresponds to partitioning with respect to the observations at hand, such as data from specific time points and sites. Hereafter, we refer to them as *sample variance partitions,* since they describe the properties of a finite sample from some larger domain, for example a set of potentially randomized sampling locations in a larger landscape or sampling times of some continuous process. In some cases, it will be possible to partition variance with respect to the full domain of interest (see Section 3.3). The domain can, for example, be the area and time period over which a survey was conducted (see Yuan *et al*. 2017). This *population variance partition* will often correspond closer to what is of ecological interest, especially when modelling observational data obtained without a well-designed survey. In such data, the distribution of environmental covariates in the observations does not correspond to that of the whole study region and the *sample variance partition* will not be representative. However, whether population variance partitions can be calculated or not, depends on the availability of covariate data over the domain and is thus often not easily realizable. The distinction between sample and population variance partitioning is demonstrated in the case study (section 3.3) and Fig. 2 ((a) and (b)).

The notion of variance partitioning over an existing, but only partially observed region or population, can be extended to *predictive variance partitioning* over new domains. For example, in climate change studies, we could be interested in predicting how much variation in species distribution is contributed by the environment under future climate compared to the current climate. When predicting variance contributions for a new location or time, the distribution, and hence variability, of the predictors often differs from the original study location or time, even if in terms of the range of values for individual covariates the prediction would have the character of "interpolation". Hence, population variance partitioning can also be a tool for summarizing ecological predictions and scenarios.

2.5 Conditional variance partitioning and partitioning between and within conditions

If we group our observations according to some categorical variable or some function that assigns observations to groups based on the covariates, we can calculate variance conditional on that grouping. For example, our case study (Section 3.4) includes a covariate for the abundance of the insect herbivore's host plant. It is divided into three categories, with all observations within the same abundance category grouped together. Formally, we define a conditioning function $g(x)$ that assigns each observation into one of G predefined groups. The variance partitioning conditional on group $c \in G$ is then calculated over the subset of observations (or the subset of the population) for which $g(x) = c$. For example, the conditional sample variance partitioning for the covariate x_j would be

$$V_{x_j|g=c}(\theta) = \frac{\mathrm{Var}\left[x_j\beta_j|g = c\right]}{\mathrm{Var}[\eta|g = c]} \tag{7}$$

where $Var[\cdot\,|g = c]$ denotes the sample variance over the subset of observations for which $g(x_i) = c$ ($i = 1, \ldots, n$ indexes the observations, see also Mathematical supplement 2.4). The definition for $P_{x_j|g=c}(\theta)$ would arise analogously.

When, for example, a predictor varies only over some subset of the observations, conditional variance partitioning can be used to study its role in that subset where it varies. However, conditional partitioning should be used with some meaningful question in mind, not just to search for an arbitrary division of the data to make up for the minor role of a predictor for the whole data. Unlike the groups of linear terms, the conditioning variable is a property of the data – that is the group membership of each observation depends on the values of the covariates (or functions thereof) available for the observations. Usually these variables would be part of the covariates of the linear predictor. Nonetheless, the choice of the conditioning variable is still made by the analyst.

We can extend conditional variance partitioning to see how much of the total variation is partitioned within the groups defined by conditioning variable and how much variation is due to the differences between the conditions. This approach has the character of analysis of variance (*cf.* Pedhazur 1997, pp. 679–683, Nakagawa and Schielzeth 2013). It utilizes the law of total (co)variance $\mathrm{Var}[x_j\beta_j] = \mathrm{E}\left[\mathrm{Var}[x_j\beta_j|g]\right] + \mathrm{Var}\left[\mathrm{E}[x_j\beta_j|g]\right]$ to split total variation into variation within ($\mathrm{E}\left[\mathrm{Var}[x_j\beta_j|g]\right]$) and between ($\mathrm{Var}\left[\mathrm{E}[x_j\beta_j|g]\right]$) conditions (Whittaker 1990, pp. 124–125).

Specifically, in applied variance partitioning we utilize the sample covariance of the linear terms within and between conditions. The variance partition within conditions is simply the weighted average of the conditional variances

$$V_{x_j|g}^{\mathrm{wi}} = \frac{\sum_{c=1}^{C}(n_{g=c} - 1)\mathrm{Var}[x_j\beta_j|g = c]}{(n-1)\mathrm{Var}[\eta]} \tag{8}$$

where $n_{g=c}$ is the number of observations belonging to the condition c and $n = \sum_{c=1}^{C} n_{g=c}$ is the total number of observations.The variation between groups is derived from the weighted variance of the per-condition means of the linear terms

$$V_{x_j|g}^{\text{btw}} = \frac{\sum_{c=1}^{C} n_{g=c}\left(\bar{x}_{j|g=c} - \bar{x}_j\right)^2}{(n-1)\text{Var}[\eta]} \tag{9}$$

where the $\bar{x}_{j|g=c}$ is the per-condition mean and $\bar{x}_j$ is the unconditional or grand mean of the linear term over all n observations. The within and between condition diagonal variance partitions $P_{x_j|g}^{\text{wi}}$ and $P_{x_j|g}^{\text{btw}}$ are defined analogously. Note that the sum of the variance partitions within and between conditions equals the variance partitioning over all observations, i.e. $V_{x_j} = V_{x_j|g}^{\text{wi}} + V_{x_j|g}^{\text{btw}}$. For derivation of the within and between condition variances see the Mathematical supplement 2.5.

Differences between the conditional variance partitions imply that the distributions of covariates and random effects differ between the conditions. The partitioning of variation to within and between condition (co)variance is a way to summarize these differences for individual linear terms or predictor groups. When the conditioning variable is a categorical predictor or the level of an *iid* random effect, then the corresponding linear term will contribute only to between-condition variation. At the other extreme, a predictor that is constrained to sum to zero within each condition would only contribute to within-condition variation.

2.6 Correlation between covariates and posterior correlation between parameters

The variance partitions $p(V|D)$ and $p(P|D)$ typically include correlations between their components, the V_xs and P_xs (see also Fig. 3) – an issue which is often overlooked when dealing only with point estimates. These correlations are a function of both correlations among the

covariates and the posterior correlations among the model parameters θ. Generally, whenever one variance component has a higher contribution, it must be offset by a reduction in the other components. How this reduction is distributed among the other components depends on the correlations between the covariates and model parameters. The correlation between variance components contributes to uncertainty in variance partitioning; that is, the role of any single component cannot be independently assessed in the presence of such correlations, just as with the covariate effects themselves. Only when the correlations are low is it always safe to work directly with the marginal distributions of the individual variance partitions. As noted above, grouping of linear terms often ameliorates this issue, as the within-group correlations will be subsumed by the marginal distribution of the grouped variance partitions (V_b; Section 2.3). Since this does not totally remove the problem, it is recommendable to explicitly report and discuss the correlations between different variance components (see Section 3.6).

The problem caused by the correlations is analogous to an over a century old problem of how to interpret variance partitions in a regression context in the presence of correlations between the model's covariates (Wright 1921, Burks 1926, Cox and Snell 1974, Bring 1995, Graham 2003, Dormann *et al*. 2013). Hence, by exploiting this literature, in addition to the measures V_x and P_x presented above, we present two other measures which can be used to summarize the covariance relationships between the linear terms and can also be interpreted as variance partitions themselves (Table 2). To simplify the notation we denote here the linear terms by $a_j = x_j\beta_j$ such that $\eta = \sum_{j=1}^{d} a_j$. We begin by rewriting the variance of the linear predictor in two alternative ways

$$\mathrm{Var}[\eta] = \sum_{j=1}^{d} \overbrace{\left(\mathrm{Var}[a_j] + \sum_{j' \neq j}^{d} \mathrm{Cov}[a_j, a_{j'}] \right)}^{\text{"marginal" variance of a predictor}} \tag{10}$$

$$\mathrm{Var}[\eta] = \mathrm{Var}[\hat{\eta}(a_1, \dots, a_d)] = \mathrm{Var}\left[\hat{\eta}\left(a_{\setminus j}\right)\right] + \overbrace{\mathrm{Var}\left[\hat{\eta}\left(a_j - \hat{a}_j\left(a_{\setminus j}\right)\right)\right]}^{\substack{\text{squared semi-partial correlation} \\ \text{between } \eta \text{ and } a_j}} \tag{11}$$

where $a_{\setminus j}$ collects all other linear terms except a_j and the hat-notation denotes linear least squares predictor so that, for example, $\hat{\eta}\left(a_{\setminus j}\right)$ is the linear least squares prediction for η using $a_{\setminus j}$; since $\eta = \sum_{j=1}^{d} a_j$, it holds that $\hat{\eta}\left(a_1, \dots, a_j\right) = \eta$. The first equality above arises by just rearranging the covariance terms in (5) and the second equality is proved for example by Whittaker (1990, pp. 134–138). In linear regression, the sum of the terms inside parenthesis on the right hand side of (10) we call the marginal variance of the corresponding linear term since it is a measure of the contribution of a linear term, while accounting for covariances with all the other linear terms (Pratt 1987, Chase 1960, Bring 1995). The squared semi-partial correlation coefficient between the linear term, a_j, and the linear predictor, η, in (11) measures the variance contribution of the linear term, that cannot be accounted for by a linear combination of the other linear terms (Whittaker 1990, pp. 134–137, Legendre and Legendre 2012, pp. 173–180). Hence, squared semi-partial correlation coefficient between a linear term and a linear predictor can be seen as the unique variance from that linear term.

Motivated by the above considerations we define the *marginal variance partition* as

$$M_{x_j} = \frac{\sum_{l'=1}^{\mathrm{m}} \mathrm{Cov}\left[a_j, a_{j'}\right]}{\mathrm{Var}[\eta]}$$

and the *unique variance partition* as

$$U_{x_j} = \frac{\mathrm{Var}\left[\hat{\eta}\left(a_j - a_j\left(a_{\setminus j}\right)\right)\right]}{\mathrm{Var}[\eta]}.$$

The marginal and unique variance partitions, M_b and U_b, for a group of linear terms are calculated analogously. The posterior distribution for both of these measures is remarkably simple to calculate by exploiting the properties of the sample covariances between the linear predictor and individual predictive terms and is easy to extend to groups of linear terms as will be illustrated in Section 3.6 and described in detail in the Mathematical supplement(2.6).

The unique variance partition is equal to V_x only in the absence of correlations between the linear terms. In all other cases, its value is less than the value of V_x and it is zero if the linear term under investigation can perfectly be explained by the other linear terms. Hence, the difference between U_x and V_x is a measure of the effect of correlations between linear terms to the variance partition (see also Burks 1926, Whittaker 1990, pp. 137–140). U_x on its own can be thought of as a lower bound of variability associated with that linear term. In the case of treatment effects in an experimental context, it would be the part of variation in the outcome that is unequivocally attributable to the treatment corresponding to that linear term (i.e., the direct effect). In the case of auxiliary variables or observational data, it simply implies that that proportion of variation in the linear term is not confounded with the other linear terms in the model. However that "unique variation" could still be confounded with covariates missing from the model. We note also that, even though in some special cases the semi-partial correlation can be interpreted as a causal effect (*cf.* Burks 1926), that is not generally the case and, hence, we should not, in general, give causal interpretation for the unique variance partition.

Unlike the unique variance partition, the marginal variance partition can give information about the effect and sign of covariances with other linear terms. The unscaled measure, $\sum_{j'=1}^{\mathrm{d}} \mathrm{Cov}\left[x_{j'}\beta_{j'}, x_j\beta_j\right]$, is also equal to the covariance between the linear term and the linear predictor (see Mathematical supplement 2.6.3). Hence, when the measure is close to zero, that linear term has no direct effect on the linear predictor, only an indirect effect via its correlations with the other terms. That indirect effect is sometimes interpreted as improving the model due to

removal of "noise" from other covariates (Bring 1995). At the same time, the marginal variance partition can also equal the variance partition V_x even in the presence of covariances, if they "cancel out". The marginal variance partition shares some properties with commonality analysis, a widely used method of variance partitioning in ecology (Legendre and Legendre 2012, 570–583), namely it sums up to one, represents the direct effect of the linear term and comprises the unique and shared variance contributions of that linear term (Chase 1960, see also Beaton 1973, Mathematical supplement 2.6.4).

We illustrate possible conclusions from comparing different values of these three measures for a single linear term or a group of linear terms in Table 3 and present some of them here: When the marginal variance partition M_x is larger than the variance partition V_x, this implies that the covariances between linear term x and the other linear terms are positive on average and increase the variability of the linear predictor η. When M_x is smaller than V_x, the covariances are negative on average. If M_x is still positive there is some confounding between the linear terms, but the linear term still contributes to variability in the linear predictor also independently. When M_x is close to zero, although V_x is not, the linear term x has no direct effect on the linear predictor, but all variance attributed to x can be explained by other covariates. When negative, the linear term acts to supress variability in the linear predictor. All the previous cases imply that the unique variance partition U_x is smaller than V_x. Lastly, when the variance partition V_x is equal to M_x even though U_x is smaller than V_x, then the effect of the covariances averages out. To summarize, the difference between V_x and U_x serves as an indicator for the presence of covariances, and hence confounding, between the linear terms, while the difference between V_x and M_x shows whether the covariances increase or suppress the variability of the linear term or the linear predictor.

2.7 Variance partition measures revisited

For any variance partitioning measure there are many properties that would be desirable (Pratt 1987, Bring 1995, Nathans *et al*. 2012). We consider here only four of them (Table 2): the variance partitions should sum to one, their value should lie between zero and one, they should reflect the effect of the covariances and when the linear terms are uncorrelated they should all equal the variance partition V_x. None of the measures presented here fulfill all of them unless the linear terms are uncorrelated (Table 2), which is why it can be useful to report variance partitions in terms of more than one measure (*cf*. Pratt 1987, Whittaker 1990, p. 130, Bring 1995). There are many other measures for variance partitioning that attempt to account for covariances to be found in the ecological and statistical literature. Some of the measures are simply reformulations of V_x, U_x, and M_x. Others, which seemingly fulfill all of the properties mentioned above, have the unfortunate feature that they do not distinguish between direct and indirect variance contributions and hence do also not indicate the presence of confounding or suppression. Hence, they cannot be said to reflect the effect of covariances in an interpretable manner. Yet others, like commonality analysis, partition variance in to an exponentially increasing number of components as a function of linear terms or are not meaningfully defined for groups of linear terms of non-linear effects, which limit their application in practice. Thus we find that U_x and M_x are the most useful ones to study the effect of covariances as they are sufficient to answer elementary questions such as do correlations between the linear terms affect the variance partition (comparison of U_x and V_x) and what is the direction and effect of the consequent covariances (comparison of U_x, M_x and V_x together).

3 Variance partitioning in practice: Case study on the Glanville fritillary butterfly

We demonstrate the application of variance partitioning using a simple species distribution model to explore factors contributing to patch occupancy in a Glanville fritillary butterfly

metapopulation in the Åland Islands, SW Finland (Fig. 1). We flesh out the methodological details of the variance partitioning by concisely explaining practical application of the theory to this case study. We demonstrate the analysis steps, present the results, and discuss them from the ecological interpretation point of view in this section. In Section 4 we provide thorough discussion on the findings from the methodological point of view.

3.1 Study system and species distribution model

The study system consists of approximately some four thousand potential habitat patches in which the butterflies' larval host plants, *Plantago lanceolata* and *Veronica spicata*, occur. These habitat patches range from dry meadows, pastures, and rocky outcroppings to roadsides and yards. The butterfly, *Melitaea cinxia*, has a single generation at this northern latitude of its range; the adult butterflies emerge in early summer and the larvae, that live greagariously most of their lives, build silken nests for overwintering. The nests are easily distinguishable and are counted during an annual autumn survey. At the same time data about the habitat is gathered, such as measures of the abundance of the host plant and indicators of grazing (Ojanen *et al*. 2013).

The model and data are simplified from our previous study analysing the relationship between occupancy and abundance in this long-term ecological study system (Schulz *et al.* 2019, 2020). The data consists of over 60 000 observations from the potential habitat patches over a span of 19 years. In addition to the nest counts, the annual records include five measures of habitat quality (host plant *vegetation*, proportion of *dry vegetation*, *grazing intensity*, *grazing presence*, presence of *both host* plants), while the effects of metapopulation structure is represented by patch area and an annual estimate of population connectivity (Model supplement, Schulz *et al*. 2020). For our analyses, we only select a subset of the data available in order to mimic a scenario where only a part of the potential study domain is surveyed (see Fig. 1). This *core area* has the highest occupancy rates (average within 17.5 % *vs*. 14.5 % outside) and comprises $n_p = 1259$ of the total 4379 habitat patches in the data set. When we refer to unobserved patches or locations

below we mean the patches outside this *core area*. In reality, the full data set contains observations also from those patches. Note that we use previous occupancy as a covariate in our model and it will be missing for cases when the corresponding habitat patch was not surveyed in the previous year. For demonstration purposes we ignore the problem and simply omit the observations for which there is no corresponding observation from the same patch in the previous year. We also exclude observations with no host plant present (n = 635) as it is really an indication that the patch is no longer habitat for the butterfly.

The occupancy state of the ith habitat patch observation is defined as $y_i = I(N_i > 0)$, where N_i is the number of winter nests found. The vector of occupancy states $y = [y_i]_{i=1}^{n}$, where $n = 20086$, collects all observations from the $n_p = 1259$ individual patches and $T = 19$ years. We model the occupancy state using logistic regression with the logit link-function such that $\mathrm{E}[y_i] = \mathrm{logit}^{-1}(\eta_i)$, where η is the linear predictor vector with elements

$$\eta_i = \alpha + \overbrace{\beta_{B_N} x_{o,i}}^{\text{Occupancy } (b_N)} + \overbrace{\sum_{j \in B_M} \beta_j x_{j,i}}^{\text{Metapopulation } (b_M)} + \overbrace{\sum_{j \in B_H} \beta_j x_{j,i}}^{\text{Habitat quality } (b_H)} + \overbrace{u(i) + v(i) + z(i)}^{\text{Random effects } (b_R)} \tag{12}$$

where α is the constant of regression, $x_{o,i} \in \{0,1\}$ is the occupancy status of the patch in the previous year (0 representing unoccupied) and *occupancy*, *metapopulation*, and *habitat quality* index sets B_N, B_M and B_H collect the "fixed effects" into covariate predictor groups (see section 2.3; note that the set B_N contains only one covariate index x_o; Model supplement). The *random effects* are collected into an additional random effects group labeled R. We use this thematic grouping of the linear terms to study which general factors affect our study system rather than individual covariates or random effects. u and v are per patch and year *iid* random effects and z is a spatio-temporal random effect.

Throughout this running example, we assume that posterior inference is implemented so that we can produce random samples from the joint posterior distribution of the model parameters and random effects. This is the most common approach to implement Bayesian inference and allows using Monte Carlo techniques in calculating the variance partition measures. Hence, we assume the posterior distribution of model parameters are represented by vectors, such as $[\alpha_1, \dots, \alpha_Q]$ where Q is the number of posterior samples. The posterior distributions of random effects are represented by matrices, such as $[u_1, \dots, u_Q]$ where u_q is a column vector containing the qth posterior sample for all n_P observation locations. In our particular model, posterior inference was performed with R-INLA (Rue *et al*. 2009, Lindgren *et al*. 2011, Lindgren and Rue 2015; INLA version 21.02.23, R version 4.0.4) and is described in detail in the Model supplement where we also report the posterior distributions of the model's parameters. For code examples using a simpler model see the Programming supplement.

3.2 Sample variance partitioning for the regression model

We are interested in partitioning the variance of the linear predictor η into the groups of linear terms indicated by brackets above the model (Eq. (12)). That is, we partition the variance of linear predictor with respect to these four groups

$$\begin{aligned}\mathrm{Var}[\eta] = \mathrm{Var}[b_N] \\ &+ \mathrm{Var}[b_M] \\ &+ \mathrm{Var}[b_H] + \mathrm{Var}[b_R] + \sum_{l \in B} \sum_{l' \in B} \mathrm{Cov}[b_l, b_{l'}] I(l \neq l')\end{aligned} \tag{13}$$

where $B = \{N, M, H, R\}$ collects the labels for the predictor groups. For example, $b_M = \sum_{j \in B_M} \beta_j x_j$, and the indicator function I ensures that the (diagonal) variances are not counted twice. The variance partition measures are then

$$V_N(\theta) = \frac{\mathrm{Var}\,[b_N]}{\mathrm{Var}[\eta]} \text{ and } P_N(\theta) = \frac{\mathrm{Var}\,[b_N]}{\mathrm{Var}[b_N]+\mathrm{Var}[b_M]+\mathrm{Var}[b_H]+\mathrm{Var}\,[b_R]}$$

for occupancy and analogous for the other groups of linear terms (see also section 2.3 above).

To construct the posterior distribution of the sample variance partition we must first construct posterior samples of the predictor group effects. The samples for the fixed effects groups are specified as, $b_{N,q} = \beta_{B_N,q} x_o$, $b_{M,q} = \sum_{j \in B_M} \beta_{j,q} x_j$, $b_{H,q} = \sum_{j \in B_H} \beta_{j,q} x_j$ where $q = 1, \dots, Q$ is the index of a Monte Carlo sample from the joint posterior distribution of all parameters and random effects and each b is a column vector with rows for each observation. A posterior sample for the random effects group is correspondingly constructed as $b_{R,q} = u_q + v_q + z_q$. Thus the linear predictor is $\eta_q = b_{N,q} + b_{M,q} + b_{H,q} + b_{R,q}$. Posterior samples for the variance partitions are then acquired as a function of the sample variance of the group of linear terms as

$$V_{N,q} = \frac{\mathrm{Var}[b_{N,q}]}{\mathrm{Var}[\eta_q]}, P_{N,q} = \frac{\mathrm{Var}[b_{N,q}]}{\mathrm{Var}[b_{N,q}] + \mathrm{Var}[b_{M,q}] + \mathrm{Var}[b_{H,q}] + \mathrm{Var}[b_{R,q}]}$$

and $\hat{V}_N$ and $\hat{P}_N$ are then simply the mean over $V_{N,1}, \dots V_{N,Q}$ and $P_{N,1}, \dots P_{N,Q}$. The approach for the other predictor groups is identical.

The biased variance partitioning is calculated by first calculating posterior means of the predictor group effects: $\hat{b}_N = \hat{\beta}_{B_N} x_o$, $\hat{b}_M = \sum_{j \in B_M} \hat{\beta}_j x_j$, and $\hat{b}_{H,j} = \sum_{j \in B_H} \hat{\beta}_j x_j$ where $\hat{\beta}_j$ is the mean over $\beta_{j,1}, \dots, \beta_{j,Q}$. The posterior mean of the random effects group is constructed analogously. After this we plug in $\hat{b}_N, \hat{b}_M, \hat{b}_H, \hat{b}_R$ into the above equations to calculate $V(\hat{\theta})$ and $P(\hat{\theta})$.

The posterior distributions of all variance partitioning measures (e.g., $p(V_M|D)$) exhibit considerable uncertainty (Fig. 2 (a)). The posterior means ($\hat{V}$ and $\hat{P}$) and posterior variances ($\mathrm{Var}[V|D]$ and $\mathrm{Var}[P|D]$) of the variance partitioning measures summarize well the location and the spread of the variance partitions whereas the biased variance partitionings ($V(\hat{\theta})$ and $P(\hat{\theta})$) are considerably biased (Fig. 2 (a)). Moreover, compared to the posterior distribution of

variance partitioning measures the biased variance partitionings ($V(\hat{\theta})$ and $P(\hat{\theta})$) mistakenly suggest that the metapopulation predictor group would have higher contribution to total variance than the random effects predictor group. These differences in the summary measures illustrate the importance of reporting the variance partitioning based on the posterior distribution. The variance partition measures (V) are lower than the corresponding diagonal variance partitioning measures (P) indicating the presence of overall positive correlations among the predictor groups. The measures are qualitatively equivalent in that they preserve the ordering of the variances of the groups of linear terms, but the posterior distribution of their ratio, $p\left(\frac{P}{V}\right)$, varies as a function of what proportion of variation in the linear predictor is due to the covariances.

The random effects, metapopulation covariates and the habitat quality covariates all contribute a large proportion to the total variation in occupancy in the core study area ($\hat{V}_R = 0.32, \hat{V}_M = 0.26, \hat{V}_H = 0.21$, Fig. 2 (a)). The role of occupancy in the previous year in the variation of current occupancy is very low ($\hat{V}_N = 0.02$). Partly this is due to expected correlations with the metapopulation covariates connectivity ($r = 0.48$) and area ($r = 0.27$) as well as host plant abundance ($r = 0.22$). Given the high rate of extinctions and colonizations previous occupancy alone is not a very good predictor of current occupancy in this system; and connectivity also is a predictor for colonizations while area as well as host plant abundance are good surrogates for population size, which does predict extinction risk (Schulz *et al*. 2020). In Section 3.6 we will take a more formal look at the role of these correlations.

3.3 Population variance partitioning

In our case study, model inference is performed only on a subset of the study region (Section 3.1, Fig. 2). However, we are interested in the relative importance of the groups of linear terms over the whole distribution range of Glanville fritillary in the Åland Islands. The locations, where data have been collected, correspond to the sample and all the locations within the distribution range of Glanville fritillary correspond to the population in our study (section 2.4). To calculate the

population variance partition, we collect into $\tilde{x}$ the covariate observations from the unsampled patches.

In order to construct samples from the posterior distribution of the population variance partitioning, we first construct posterior predictive samples for the predictor groups at unobserved patches. For fixed effects groups they are specified analogously to sample variance partitioning (section 3.2) as column vectors $\tilde{b}_{N,q} = \beta_{B_N,q}\tilde{x}_o$, $\tilde{b}_{M,q} = \sum_{j\in B_M} \beta_{j,q}\tilde{x}_{M_j}$, $\tilde{b}_{H,q} = \sum_{j\in B_H} \beta_{j,q}\tilde{x}_{H_j}$ where $q = 1, \dots, Q$ is the index of a Monte Carlo sample from the joint posterior distribution of all parameters and random effects. For the random effects' predictor group $\tilde{b}_{R,q} = \tilde{u}_q + \tilde{v}_q + \tilde{z}_q$ we need to construct the predicted random effects at the unobserved locations. In our case, the sample data cover the same years as the predictions, so the yearly random effects are simply $\tilde{v}_q = v_q$. For the patch effects we must sample for each unobserved habitat patch new values from a univariate normal distribution with variance corresponding to the qth posterior sample of the variance parameter of the patch random effect. The case for the spatio-temporal random effect the approach is similar, but more complicated due to the covariance structure of the spatio-temporal random field which depends both on the hyperparameters defining the covariance matrix as well as the existing random effects z_q at the observed patches. For technical details see the Model supplement.

Once we have constructed the samples for predictor group predictors at the unobserved locations, for each sample we concatenate the unobserved predictor group column vector with the corresponding column vectors at the observed locations to form, $b_{N,q}^{\text{pop}} = \left[\tilde{b}_{N,q}^T, b_{N,q}^T\right]^T$ and equivalently for the other predictor groups. After this, we can proceed as before and calculate the *population variance partition* over the whole study region as

$$V_{N,q}^{\text{pop}} = \frac{\text{Var}[b_{N,q}^{\text{pop}}]}{\text{Var}\left[\eta_q^{\text{pop}}\right]}, P_{N,q}^{\text{pop}} = \frac{\text{Var}[b_{N,q}^{\text{pop}}]}{\text{Var}\left[b_{N,q}^{\text{pop}}\right] + \text{Var}\left[b_{M,q}^{\text{pop}}\right] + \text{Var}\left[b_{H,q}^{\text{pop}}\right] + \text{Var}\left[b_{R,q}^{\text{pop}}\right]}$$

where $\eta_q^{\text{pop}} = b_{N,q}^{\text{pop}} + b_{M,q}^{\text{pop}} + b_{H,q}^{\text{pop}} + b_{R,q}^{\text{pop}}$. A variance partitioning calculated with only $\tilde{b}_q$, such as $\tilde{V}_{\text{q}} = \frac{\text{Var}[\tilde{b}_q]}{\text{Var}[\tilde{\eta}_q]}$ would correspond to *predictive variance partitioning* (section 2.4) over the unobserved locations. In principle, it would be straightforward to replace the predictive covariate values, $\tilde{x}$ and $\tilde{y}$, with some scenario-based covariates, such as covariates corresponding to climate or habitat change.

The population variance partitioning differs from the sample variance partitioning considerably in terms of the role of the metapopulation predictors (Fig. 2 (a) and (b)). Over the whole habitat range in the Åland Islands, it accounts for approximately half of the variation in the occupancy of the butterfly, with correspondingly reduced roles for the random effects and habitat quality ($\hat{V}_R^{\text{pop}} = 0.21, \hat{V}_M^{\text{pop}} = 0.50, \hat{V}_H^{\text{pop}} = 0.14$).

This suggest that outside of the core study area connectivity and habitat patch size determine variation in occupancy as there is less variation in habitat quality ($\tilde{V}_H = 0.13$ *vs*. $\hat{V}_H = 0.21$). This is consistent with the fact that the core study area contains most of the more viable habitat of the butterfly in the Åland islands (Hanski *et al*. 2017). Also, the habitat patches outside of the core area tend to be more isolated accentuating the role of connectivity in predicting occupancy.

3.4 Variance partitioning conditional on host plant abundance

The amount of host plant vegetation within the habitat patch is known to be crucial for the specialist larvae of the butterfly (Opedal *et al*. 2020). We thus want to know what other factors may contribute to variation in occupancy after we have taken the amount of host plant vegetation into account. For this we partition variation conditional on the value of the host plant covariate (vegetation), x_v, which conveniently for our demonstration is measured on an ordinal scale from 1 to 3 with 0 indicating absence of the host plant. We define the conditioning function $g(x) = x_V$. The conditional variance partition is calculated just as the regular, but separately within the

observations belonging to each value $c = 1,2,3$ of the vegetation covariate x_v (see section 2.5). The variance partition for occupancy conditional on the vegetation covariate is

$$V_{N|x_v=c,q} = \frac{\mathrm{Var}[b_{N,q}|x_v = c]}{\mathrm{Var}[\eta_q|x_v = c]}, c \in \{1,2,3\}$$

and is calculated equivalently for the other predictor groups.

Differences between the conditional variance would imply that the other covariates or the random effects covary with host plant abundance. Or more precisely, that the distribution of the covariates (and random effects) differs between the levels of host plant abundance. Here, we see mostly only small differences between the abundance classes except for the random effects, where less variation is associated with random effects in the highest abundance class (Fig. 2 (c)). The minor role of the habitat quality predictor group indicates that the predictor group's variance contribution is mostly due to the conditioning variable host plant abundance. Again the posterior mean and variance estimates (e.g. $\hat{V}$ and $\mathrm{Var}[V|D]$) summarize well the posterior distributions for variance partitioning. However, especially for the random effects, the biased point estimates greatly underestimate the conditional variance proportions.

3.5 Variance within and between patches

We go one step further and ask whether more variation in occupancy is due to differences between patches or due to annual variation within patches and which factors drive these differences (see also section 2.5). To do this, we define the conditioning function to be $g(x) = x_P$ where x_P is an indicator vector of length n for $n_P = 1259$ patches, with $x_{P,i} = p(i)$ above.

The variation within patches $V^{\mathrm{wi}}_{x|x_P}$ is a weighted average of their conditional variances, while the variance between patches $V^{\mathrm{btw}}_{x|x_P}$ is the weighted variance of the patch means (averaged over time) of the predictor group effects. The weighing is necessary to allow for unbalanced group sizes, in

this case differing numbers of observations per patch. A posterior sample from the within and between condition variance for occupancy are formed as

$$V_{N|x_P,q}^{\text{wi}} = \frac{\sum_{c=1}^{n_P}\left(n_{x_P=c}-1\right)V_{N|x_P=c,q}}{n-1}$$

$$V_{N|x_P,q}^{\text{btw}} = \frac{\sum_{c=1}^{n_P} n_{x_P=c}\left(\bar{b}_{N|x_P=c,q}-\bar{b}_{N,q}\right)^2}{(n-1)\text{Var}\left[\eta_q\right]}$$

Here $\bar{b}_{N,q}$ is the mean of occupancy over all observations, while $\bar{b}_{N|x_P=c,q}$ is the per-patch mean for occupancy at patch c both calculated with the qth Monte Carlo sample of parameters. That is $\bar{b}_{N,q} = \frac{1}{n}\sum_{i=1}^{n}\left[b_{N,q}\right]_i$ and, $\bar{b}_{N|x_P=c,q} = \frac{1}{n_{x_P}}\sum_{i\in x_P=c}\beta_{N,q}y_{t(i)-1,p(i)}$. The calculation for the other predictor groups follow the same pattern.

When the variance contributions are separated for within and between habitat patches we see that the metapopulation covariates contribute more to differences in occupancy between patches than interannual variability within the patches (Fig. 2 (d)). This is to be expected as connectivity varies significantly between patches already simply due to varying isolation, though annual variation is also present (Hanski *et al*. 2017, Kahilainen *et al*. 2018). And as the data uses a single areal measurement for each patch, habitat area cannot contribute at all to within patch variation. For the random effects the situation is reversed and they contribute somewhat more to yearly variation in occupancy within the patches. The habitat quality covariates contribute almost equally to variation within and between patches. This dual role for habitat quality is consistent with the observation that in addition to habitat patches vastly differing in quality (Schulz *et al*. 2020) there is also significant annual variation in host plant availability within these patches (Opedal *et al*. 2020).

The biased point estimates are rather close to the actual variance partitions in all cases but one. The estimate for the random effects' contribution to variation between patches is severely

underestimated. A likely explanation lies in the fact that the posterior mean of the random effects is close to zero, which is both a function of the random effects' priors having a mean of zero and an constraint for the random effects to sum to zero (Model supplement). As a consequence a large proportion of their variance contribution is due to the posterior variance, which is ignored by the biased point estimate. This is a general problem whenever the mean estimate is close to zero, but the uncertainty about the estimate is great, which almost by design is true of random effects, but can affect covariate terms as well.

3.6 Correlation between covariates and posterior correlation between parameters

To understand the effects of correlated covariates and random effects, we start by visualizing both the posterior distributions of the covariances between predictor groups and the correlations of the predictor group's variance partitions (Fig. 3). The unique variance partition for the groups of linear effects is the proportion of variation due to the linear terms in that group, which cannot be expressed as a linear combination of linear terms belonging to other groups. The measure can calculated using the sample covariance matrix of the linear terms (Mathematical supplement 2.6.2):

$$U_{N,q} = K_{N,N-} K_{N,\backslash N} K_{\backslash N}^{-1} K_{\backslash N,N}$$

where K is the sample covariance matrix between the linear terms in the model, N indexes the rows or columns belonging to occupancy and $\backslash N$ all the linear terms in the other groups of linear terms. Note that in general U_b is the sum of the variances in the above equation, but that is omitted as the occupancy group consists of a single covariate.

The marginal variance partitions for groups of linear terms, M_b, are calculated as the variance partition plus the covariance with the other groups:

$$M_{N,q} = V_{N,q} + \frac{\mathrm{Cov}[b_{N,q}, b_{R,q}] + \mathrm{Cov}[b_{N,q}, b_{M,q}] + \mathrm{Cov}[b_{N,q}, b_{H,q}]}{\mathrm{Var}[\eta_q]}$$

Some of the groups of linear terms with the high correlations between them have also high posterior correlations between their variance measures (see Fig. 3). But looking at correlations of pairs of linear term groups alone is not always sufficient to surmise the posterior correlations between their variance partitions. Rather, the correlations between the variance partitions posterior distributions are due to both the correlations among the covariates and the posterior correlations among the model parameters. Even if all covariates were mutually orthogonal, there could be posterior correlation between the variance partitions due to posterior correlations between model parameters. The effect of covariate correlations on the posterior correlation between two linear terms depends not only on the correlation between the corresponding covariates, but the size of their variance contributions and the correlations between all the other covariates. The resulting correlations are not always intuitive. As a stark example of this, the random effects and habitat quality covariates have a very low correlation, with the posterior distribution of covariances overlapping zero, yet their variance partitions have strong negative correlations ($\mathrm{Cor}[V_R(\theta), V_H(\theta)] = -0.33$ by expectation). The same is true for the random effects and the metapopulation covariates ($\mathrm{Cor}[V_R(\theta), V_M(\theta)] = -0.23$ by expectation). The posterior distribution of the random effects' variance partition is negatively correlated with all the other variance partitions (Fig. 3) precisely because its variance contribution is mostly unique (Fig . 4). Any increase in its proportion of total variation will have to be offset by the other groups' variance partitions. And the three groups of covariate linear terms are all correlated, but only the metapopulation and habitat quality variance partitions have a clear negative correlation (Fig. 3); the occupancy and the habitat quality linear term groups are clearly correlated ($\mathrm{E}\big[\mathrm{Cor}[b_N, b_H]\big] = 0.23$), but the correlation between their variance partitioning measures is close to zero.

The random' effects unique variance partition U_R is approximately equal to its variance partition V_R and the marginal variance partition M_R is correspondingly only slightly higher than V_R (Fig .

4). This means there is only minor confounding between the random effects and the covariates, even though the posterior distribution of V_R is clearly correlated with the other variance partitions. All of the marginal variance partitions are higher than their corresponding variance partitions, which means that the covariances between the groups of linear terms are overall positive. The clearest example is occupancy, whose main contribution to variability is due to its covariances with the other linear terms. This is not unexpected as due to the high extinction and colonization rates in this system occupancy alone is not a generally reliable predictor of future occupancy states and as occupancy is highly correlated with habitat quality, patch area and connectivity. It is interesting to note that the metapopulation covariates and habitat quality covariate have approximately equal unique contributions, but the variance and marginal variance partition are both higher for the metapopulation covariates. A potential explanation is the higher correlation between occupancy and metapopulation covariates than occupancy and habitat quality covariates (Fig. 3).

4 Discussion

Variance partitioning in its many forms has been an important tool in ecology already for decades and we believe it continues to hold its central role in future as well. As often happens with highly influential methods, variance partitioning has evolved to many directions over the years. Some of these directions, however, have lost their connection to the original theory whereas others have improved and developed the core ideas significantly further. In this work, we have discussed the pros and cons of some of the modern versions of variance partitioning and proposed a few extensions to them that we see useful. Our work is not intended to be a review on the subject but we have taken a practical stance by pointing our criticism to popular approaches that we think are problematic while appraising approaches that we think should be more widely adopted in ecological research. We treat the subject from Bayesian perspective since it allows transparent and straightforward treatment of uncertainties.

The central idea in variance partitioning is to examine the relative importance of alternative ecological processes. As shown in our earlier presentation of the methods and exemplified in by our case study, the form of variance partitioning should change according to the question asked. For example, the sample variance partitioning for individual model components (the individual covariates in our case study) informs us about the relative roles of those covariates only (Section 2.2). However, distinct covariates often fall into broader sets of ecological processes, such as occupancy, metapopulation, habitat and random processes in our case study. If we are interested in the relative importance of these higher level processes in our study system, we should decrease the resolution of the analysis by using variance partitioning for the groups of linear terms formed by these upper level processes (Sections 2.3, 3.2). At the same time, the conditional variance partitioning and its extension to between and within condition variance partitioning increases the resolution of answers available by allowing the analyst to zoom into different compartments of the study system. That provides a method to assess what are the relative roles of the ecological processes in the subset of cases falling into a particular compartment, and whether the relative roles of the ecological processes are consistent between the compartments (low between condition variance) or not (high between condition variance) (Sections 2.5–2.6, 3.4–3.5).

Moreover, variance partitioning extends straightforwardly beyond the study system from where the data are collected. Here we demonstrated this by calculating the population variance partitioning for the larger system in whole from which we had (hypothetically) observed only a small fraction (section 3.3). We could analogously extend the idea to predictive variance partitioning analysis related to, for example, climate or land use change. These approaches to variance partitioning presented take a finite-population view of variation by looking at sample and population variance partitions for a specific study system. One essential question in ecology is, however, to distinguish between context dependency and generality. The variance partitioning can be turned to this question as well by implementing so called superpopulation view of

variance (Gelman *et al*. 2014, pp. 201–202). For example, we could model spatially distinct study systems (such as, Glanville fritillary populations in Åland or UK) with a hierarchical Bayesian framework by assigning variance parameters to each study system and to a higher hierarchical level description of consistent environmental response across systems. The variance parameters could be interpreted as measures of variability across a larger domain than encompassed by any single study system. For specific questions such model could also make predictions for unobserved locations. In our context, this could for example involve the prediction of occupancy in populations elsewhere in the study species' distribution. Superpopulation variance partitioning requires a more carefully thought out model structure and is more sensitive to choice of priors (Gelman 2005, 2006). Developing these methods further is beyond the scope of this work.

Our approach to variance partitioning is strictly model based, which has several consequences. All these analyses can be attained by a single model after a single parameter inference step. The model based approach allows us to coherently encode the ecological understanding into analysis. It also allows, when combined with Bayesian approach to parameter inference, transparent and theoretically justified treatment of uncertainty. These properties are clearly an advantage when our model explains the system well but it is also a reminder that we should not jump into variance partitioning before scrutinizing our model.

The treatment of uncertainty is one of the central themes in our work. In our experience, this is typically neglegted at many mutually connected levels of variance partitioning. Firstly, and most evidently, the variance partitioning summaries themselves are uncertain. This simple fact seems to be forgotten in most applications of variance partitioning even if authors had careful uncertainty treatment in other parts of their analysis. As shown in our results (Fig. 2), ignoring the uncertainty in variance partitioning by using simple parameter estimates may significantly change quantitative but also the qualitative conclusions from the study compared to treating the

uncertainty correctly. Variance partitions should be treated as random variables and should be interpreted either by studying their distributions directly or by suppressing these distributions to well defined summary statistics such as posterior mean and variance or credible intervals. As shown in our case study, propagating the parameter uncertainty is rather simple in practice, as it can be done after model inference and only involves the same steps as calculating the variance partition in the first place. Hence, there is no reason to ignore the uncertainty in variance partitioning.

The above uncertainty considerations apply to the marginal distribution of variance partitions but another aspect of uncertainty is driven by the correlations between variance partitions. This is induced by correlations between model covariates and the correlations between model parameters in their posterior distribution. Even in the applied Bayesian variance partitioning studies that report uncertainties of marginal distributions, this aspect of uncertainty seems to have received barely any attention. For this reason we advocate explicitly reporting posterior correlations between variance partitions and linear terms as well as comparing marginal distribution of variance partitions to the unique variance partition and the marginal variance partition. These latter measures summarize the covariance relationships between the linear terms and, together with the basic variance partition, help interpreting possible confounding effects on the results.

When our research involves observational data, we will inevitably encounter situations where our covariates are correlated. The first step in approaching analyses of such data is acceptance. Correlations might be a nuisance from the analyst's perspective, but they are also a real feature of the ecological system with implications for how the system works. In terms of partitioning variation, the resulting covariances between the linear terms are a feature not a deficiency in the method. They summarize the effects of the covariate distribution on variability and the limits of which effects can be identified independently. Hence we recommend not to remove covariates

from an analysis simply because they are correlated (*cf*. Cox and Snell 1974, Morrisey and Ruxton 2018) as this only hides the *problem* and gives a false sense of certainty. We also do not recommend principal components analysis (PCA) as a general solution to correlated covariates. To remain meaningfully interpretable, the covariates included in PCA should be sufficiently similar. Moreover, in attempting to interpret the PCA one has not escaped the correlations after all, but moved the problem to the interpretation of the loading factors. And when working with non-linear or additive effects it does not seem sensible to use anything but the original covariates.

If the aim is to reduce the number of covariates to a smaller group of meaningful features, this can be achieved with grouping of the linear terms. Especially when interest is less in interpreting regression coefficients or prediction and more in looking at variance contributions, the problem of correlations is reduced by the grouping of the linear terms. When the correlation between covariates is strong enough that their independent effects cannot be estimated in a model, by grouping the linear terms one can still estimate their joint contribution to the model (*cf*. Mood 1971). Generally, we recommend only grouping thematically related covariates. Depending on the research question, the analyst might well be interested even in reporting variances for predictor groups consisting of rather dissimilar covariates. Either approach provides an alternative to excluding correlated covariates, a practice often arising from a misunderstanding about the effect and role of collinearity (Cox and Snell 1974, Morrisey and Ruxton 2018).

The simple fact that variances of sums equal sums of variances gets us surprisingly far. We like to think of the linear terms and their variances as the common currency of regression models. They are all on the same scale and can be meaningfully added together, which makes variance partitioning such a versatile approach — while we should not add apples and oranges, we can freely add growth from apples and oranges eaten to get total growth. Unlike working only directly with regression coefficients, partial regression coefficients, or especially non-linear

effects, variance partitioning can deal with both individual predictors and groups of them. And the within–between condition comparison is not available from simple regression coefficients which also do not account for effect size *vs*. realized variability.

Acknowledgements

We wish to acknowledge Elina Numminen and Aapo Kahilainen for their critical comments on the manuscript. TS was funded by the Doctoral Programme in Wildlife Biology Fellowship (Univ. of Helsinki) and MS received funding from the European Research Council (Independent Starting grant no. 637412 ’META-STRESS’) and JV from the Academy of Finland (grant 317255).

5 References

Beaton Jr., A. E. 1973. *Commonality*. U.S. Dept. of Health, Education, and Welfare, National Institute of Education.

Bring, J. 1994. How to standardize regression coefficients. *The American Statistician* 48: 209. doi:10.2307/2684719

Bring, J. 1995. "Variable importance by partitioning R^2." *Quality & Quantity* 29: 173–89. doi:10.1007/BF01101897

Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the spatial component of ecological variation. *Ecology* 73: 1045–55. doi:10.2307/1940179

Burks, B. S. 1926. On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology* 17: 532–540. doi:10.1037/h0070156

Casella, G., and R. L. Berger. 2002. *Statistical Inference*. 2nd ed. Pacific Grove: Duxbury.

Casini, M., M. Cardinale, and J Hjelm. 2006. Inter-annual variation in herring, *Clupea harengus*, and sprat, *Sprattus sprattus*, condition in the Central Baltic Sea: What gives the tune? *Oikos* 112: 638–50. doi:10.1111/j.0030-1299.2006.13860.x

Chase, C. I. 1960. Computation of Variance Accounted for in Multiple Correlation. *The Journal of Experimental Education* 28: 265–266. doi:10.1080/00220973.1960.11010660

Chevan, A., and M. Sutherland. 1991. Hierarchical partitioning. *The American Statistician* 45: 90–96. doi:10.1080/00031305.1991.10475776

Collie, J. S., R. M. Peterman, and B. M. Zuehlke. 2012. A fisheries risk-assessment framework to evaluate trade-offs among management options in the presence of time-varying productivity. *Canadian Journal of Fisheries and Aquatic Sciences* 69: 209–23. doi:10.1139/f2011-148.

Cooper, G. 1998. Generalizations in ecology: A philosophical taxonomy. *Biology & Philosophy* 13: 555–586. doi:10.1023/A:1006508101996

Cox, D.R., and Snell, E.J. 1974. The choice of variables in observational studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 23: 51–59. doi:10.2307/2347053

Cressie, N., and C. K. Wikle. 2011. *Statistics for spatio-temporal data*. John Wiley & Sons.

Cushman, S. A., and K. McGarigal. 2002. Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecology* 17: 637–46. doi:10.1023/A:1021571603605

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., and Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27–46. doi:10.1111/j.1600-0587.2012.07348.x

Estay, S. A., M. Lima, and F. Bozinovic. 2014. The role of temperature variability on insect performance and population dynamics in a warming world. *Oikos* 123: 131–40. doi:10.1111/j.1600-0706.2013.00607.x

Foster, S. D., J. Vanhatalo, V. M. Trenkel, T. Schulz, E. Lawrence, R. Przeslawski, and G. R. Hosack. 2021. Effects of ignoring survey design information for data reuse. *Ecological Applications*: e2360. doi:10.1002/eap.2360

Gelman, A. 2005. Analysis of variance—Why it is ore important than ever. *The Annals of Statistics* 33: 1–53. doi:10.1214/009053604000001048

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis* 1: 515–34. doi:10.1214/06-BA117A

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. 3rd ed. Chapman & Hall/CRC Texts in Statistical Science Series. Boca Raton: CRC Press.

Gelman, A., B. Goodrich, J. Gabry, and A. Vehtari. 2019. R-squared for bayesian regression models. *The American Statistician* 73: 307–309. doi:10.1080/00031305.2018.1549100

Goldstein, H., W. Browne, and J. Rasbash. 2002. Partitioning variation in multilevel models. *Understanding Statistics* 1: 223–31. doi:10.1207/S15328031US0104_02

Graham, M. H. 2008. Confronting multicollinearity in ecological multiple regression. *Ecology* 84: 2809–2815. doi:10.1890/02-3114

Hanski, I., T. Schulz, S. C. Wong, V. Ahola, A. Ruokolainen, and S. P. Ojanen. 2017. Ecological and genetic basis of metapopulation persistence of the Glanville fritillary butterfly in fragmented landscapes. *Nature Communicatios* 8: 14504. doi:10.1038/ncomms14504

Hodges, J. S., Y. Cui, D. J. Sargent, and B. P. Carlin. 2007. Smoothing balanced single-error-term analysis of bariance. *Technometrics* 49: 12–25. doi:10.1198/004017006000000408

Hubbell, S. P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology 32. Princeton: Princeton University Press, 2001.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. *Ecological Monographs* 73: 301–30. doi:10.1890/0012-9615(2003)073[0301:ECSAEI]2.0.CO;2

Janssen, F., C. Schrum, and J. O. Backhaus. 1999. A climatological data set of temperature and salinity for the Baltic Sea and the North Sea. *Deutsche Hydrographische Zeitschrift* 51: 5–245. doi:10.1007/BF02933676

Kahilainen, A., S. van Nouhuys, T. Schulz, and M. Saastamoinen. 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 Biology* 24: 4316–29. doi:10.1111/gcb.14280

Lappalainen, J., and H. Lehtonen. 1995. Year-class strength of pikeperch (*Stizostedion Lucioperca* L.) in relation to environmental factors in a shallow Baltic bay. *Annales Zoologici Fennici* 32: 411–419.

Legendre, P., and L. Legendre. 2012. *Numerical ecology*. 3rd ed. Amsterdam: Elsevier.

Levins, R., and R. Lewontin. 1982. Dialectics and reductionism in ecology. In: *Conceptual Issues in Ecology*, ed. Esa Saarinen: 107–38. Dordrecht: Springer Netherlands. doi:10.1007/978-94-009-7796-9_4

Lindgren, F. and Rue, H. 2015. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* 63: 1–25. doi:10.18637/jss.v063.i19

Lindgren, F., H. Rue, and J. Lindström. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73: 423–498. doi:10.1111/j.1467-9868.2011.00777.x

Mac Nally, R. 1996. Hierarchical partitioning as an interpretative tool in multivariate inference. *Austral Ecology* 21: 224–28. doi:10.1111/j.1442-9993.1996.tb00602.x

MacKenzie, B. R., H. Gislason, C. Möllmann, and F. W. Köster. 2007. Impact of 21st century climate change on the Baltic Sea fish community and fisheries. *Global Change Biology* 13: 1348–67. doi:10.1111/j.1365-2486.2007.01369.x

McCallum, H. 2016. Models for managing wildlife disease. *Parasitology* 143: 805–820. doi:10.1017/S0031182015000980

McMahon, S. M., and J. M. Diez. 2007. Scales of association: Hierarchical linear models and the measurement of ecological systems. *Ecology Letters* 10: 437–52. doi:10.1111/j.1461-0248.2007.01036.x

Mood A. 1971. Partitioning variance in multiple regression analyses as a tool for developing learning models. *American Educational Research Journal* 8:191–202. doi:10.3102/00028312008002191

Morrissey, M. B., and G. D. Ruxton. 2018. Multiple regression is not multiple regressions: The meaning of multiple regression and the non-problem of collinearity. *Philosophy, Theory, and Practice in Biology* 10: 3. doi: 10.3998/ptpbio.16039257.0010.003

Mulaik, S. A. 2010. *Foundations of Factor Analysis*. 2nd ed. New York: Chapman and Hall/CRC. doi:10.1201/b15851

Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effec ts models. *Methods in Ecology and Evolution* 4: 133–142. doi:j.2041-210x.2012.00261.x

Nathans, Laura L., Oswald, Frederick L., and Nimon, Kim. 2012. Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research, and Evaluation* 17: 9. doi:10.7275/5fex-b874

Nissling, A., U. Johansson, and M. Jacobsson. 2006. Effects of salinity and temperature conditions on the reproductive success of turbot (*Scophthalmus maximus*) in the Baltic Sea. *Fisheries Research* 80: 230–38. doi:10.1016/j.fishres.2006.04.005

Ojanen, S. P., M. Nieminen, E. Meyke, J. Pöyry, and I. Hanski. 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. doi:10.1002/ece3.733

Opedal, Ø. H., O. Ovaskainen, M. Saastamoinen, A.-L, Laine, and S. van Nouhuys. 2020. Host plant availability drives the spatio-temporal dynamics of interacting metapopulations across a fragmented landscape. *Ecology* 101: e03186. doi:10.1002/ecy.3186

Ovaskainen, O., H. J. de Knegt, and M. M. Delgado. 2016. *Quantitative Ecology and Evolutionary Biology: Integrating Models with Data*. Oxford University Press. doi:10.1093/acprof:oso/9780198714866.001.0001

Ovaskainen, O. and N. Abrego. 2020. *Joint Species Distribution Modelling: With Applications in R*. Ecology, Biodiversity and Conservation. Cambridge, United Kingdom ; New York, NY: Cambridge University Press.

Pedhazur, E. J. 1997. *Multiple regression in behavioral research: Explanation and prediction*. New York: Thompson Learning.

Peres-Neto, P. .R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* 87: 2614–2625. doi: 10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2

Péron, G., and D. N. Koons. 2012. Integrated modeling of communities: parasitism, competition, and demographic synchrony in sympatric ducks. *Ecology* 93: 2456–2464. doi:10.1890/11-1881.1

Pratt, J. W. 1987. Dividing the indivisible: Using simple symmetry to partition variance explained. In: *Proceedings of the 2nd International Tampere Conference in Statistics*, eds. Pukkila, T., and S. Puntanen: 245–60. Department of Mathematical Sciences, University of Tampere.

Ray-Mukherjee, J., K, Nimon, S. Mukherjee, D. W. Morris., R. Slotow, and M. Hamer. 2014. Using commonality analysis in multiple regressions: a tool to decompose regression effects in

the face of multicollinearity. *Methods in Ecology and Evolution* 5: 320–328. doi:10.1111/2041-210X.12166

Ricker, W. E. 1954. Stock and recruitment. *Journal of the Fisheries Research Board of Canada* 11: 559–623. doi:10.1139/f54-039

Riitters, K. H., R. V. O'Neill, C. T. Hunsaker, J. D. Wickham, D. H. Yankee, S. P. Timmins, K. B. Jones, and B. L. Jackson. 1995. A factor analysis of landscape pattern and structure metrics. *Landscape Ecology* 10: 23–39. doi:10.1007/BF00158551

Rue, H., S. Martino, and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71: 319–392. doi:10.1111/j.1467-9868.2008.00700.x

Ruppert, D., and R. J. Carroll. 1985. Data transformations in regression analysis with applications to stock—recruitment relationships. In: *Resource Management*, ed. Marc Mangel, 61: 29–47. Lecture Notes in Biomathematics. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-46562-8_2

Schulz, T., J. Vanhatalo, and M. Saastamoinen. 2019. Data from: Long-term demographic surveys reveal a consistent relationship between average occupancy and abundance within local populations of a butterfly metapopulation. *Dryad*. Dataset. doi:10.5061/dryad.ksn02v707

Schulz, T., J. Vanhatalo, and M. Saastamoinen. 2020. Long-term demographic surveys reveal a consistent relationship between average occupancy and abundance within local populations of a butterfly metapopulation. *Ecography* 43: 306–317. doi:10.1111/ecog.04799

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. *Variance components*. Hoboken, NJ: Wiley.

Teng, S. N., C. Xu, B. Sandel, and J.-C, Svenning. 2018. Effects of intrinsic sources of spatial autocorrelation on spatial regression modelling. *Methods in Ecology and Evolution* 9: 363–372. doi:10.1111/2041-210X.12866

Tikhonov, G., Ø. H. Opedal, N. Abrego, A. Lehikoinen, M. M. J. Jonge, J. Oksanen, and O. Ovaskainen. 2020. Joint Species Distribution Modelling with the R -Package HMSC. *Methods in Ecology and Evolution* 11: 442–47. doi:10.1111/2041-210X.13345

Tomkiewicz, J., K. M. Lehmann, and M. A. St. John. 1998. Oceanographic influences on the distribution of baltic cod, *Gadus morhua*, during spawning in the Bornholm Basin of the Baltic Sea. *Fisheries Oceanography* 7: 48–62. doi:10.1046/j.1365-2419.1998.00051.x

Vellend, M.. 2010. Conceptual synthesis in community ecology. *The Quarterly Review of Biology*: 85: 183–206. doi:10.1086/652373

Veneranta, L., L. Urho, A. Lappalainen, and M. Kallasvuo. 2011. Turbidity characterizes the reproduction areas of pikeperch (*Sander lucioperca* (L.)) in the morthern Baltic Sea. *Estuarine, Coastal and Shelf Science* 95: 199–206. doi:10.1016/j.ecss.2011.08.032

Veneranta, L., J. Vanhatalo, and L. Urho. 2016. Detailed temperature mapping–warming characterizes archipelago zones. *Estuarine, Coastal and Shelf Science* 182: 123–35. doi:10.1016/j.ecss.2016.09.011

Vernier, P. R., F. K. A. Schmiegelow, S. Hannon, Susan and S. G. Cumming. 2008. Generalizability of songbird habitat models in boreal mixedwood forests of Alberta. Ecological Modelling 211: 191–201. doi: 10.1016/j.ecolmodel.2007.09.004

Vuorinen, I., J. Hänninen, M. Rajasilta, P. Laine, J. Eklund, F. Montesino-Pouzols, F. Corona, K. Junker, H. E. M. Meier, and J. W. Dippner. 2015. Scenario simulations of future salinity and ecological consequences in the Baltic Sea and adjacent North Sea areas–Implications for

environmental monitoring. *Ecological Indicators* 50: 196–205. doi:10.1016/j.ecolind.2014.10.019

Wakefield, J. 2003. Sensitivity analyses for ecological regression. *Biometrics* 59: 9–17. doi:10.1111/1541-0420.00002

Wakefield, J. 2004. A critique of statistical aspects of ecological studies in spatial epidemiology. *Environmental and Ecological Statistics* 11: 31–54. doi:10.1023/B:EEST.0000011363.12720.38

Weigel, B., J. Mäkinen, M. Kallasvuo, and J. Vanhatalo. 2021. Exposing changing phenology of fish larvae by modeling climate effects on temporal early life-stage shifts. *Marine Ecology Progress Series* 666: 135–148. doi:10.3354/meps13676

Wisler, C. E. 1969. On partitioning the explained variation in a regression analysis. In: *A Study of our Nation's School,*. eds. Mayeske, G. W. *et al*: 344–360. Appendix II: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Whitlock, M. C., and R. Gomulkiewicz. 2005. Probability of fixation in a heterogeneous environment. *Genetics* 171: 1407–1417. doi:10.1534/genetics.104.040089

Whittaker, J. 1990. *Graphical models in applied multivariate statistics*. United Kingdom: John Wiley & Sons.

Wright, S. 1921. Correlation and causation. *Journal of Agricultural Research*: 557–585.

Yuan, Y., F. E. Bachl, F. Lindgren, D. L. Borchers, J. B. Illian, S. T. Buckland, H. Rue, and T. Gerrodette. 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics* 11: 2270–97. doi:10.1214/17-AOAS1078

Tables

Table 1 Notation and terminology for finite population variance partitioning measures.

Notation	Concept	Comment
V_x	*Variance partition* as a random variable	Proportion of variation in the linear predictor due to the linear term x.
P_x	*Diagonal variance partition* as a random variable	Proportion of independent variation in the linear term x.
$V_x(\hat{\theta})$	Variance partition at a point estimate	Biased estimate of V_x calculated from a point estimate of the parameters.
$P_x(\hat{\theta})$	Diagonal variance partition at a point estimate	Biased estimate of P_x calculated from a point estimate of the parameters.
$\hat{V}_x$	Point estimate of the variance partition	Location point estimate (e.g. posterior mean) of the distribution of V_x.
$\hat{P}_x$	Point estimate of the diagonal variance partition	Location point estimate (e.g. posterior mean) of the distribution of P_x.
$\mathrm{Var}[V_x]$	Posterior variance of the variance partition	Summary of the posterior uncertainty in the variance partition estimate.
$\mathrm{Var}[P_x]$	Posterior variance of the diagonal variance partition	Summary of the posterior uncertainty in the diagonal variance partition estimate.
b_l	*predictor group* effect in generalized linear regression	Linear combination of the linear terms belonging to the group l.
V_x	*Sample variance partition*	Variance partition calculated over the observations included in the study
V_x^{pop}	*Population variance partition*	Variance partition calculated over the population of which the observations are a sample.
V_x^{pred}	*Predictive population variance partition*	Variance partition calculated over some new, possibly hypothetical, set of observations.

Table 2. Various variance partitioning measures derived from the covariance matrix of the linear terms. The *Sum* column indicates the sum of the measure over all linear terms. The *Value* column shows the range of values the measure can take for a single linear term; an empty cell means there is no constraint on the value. An "x" in the *Covariance* column indicates that the measure is a function of the covariance between the linear terms. Note that in the context of model inference, even P_x is of course affected by the covariance but the variance measure itself can be calculated without referencing the covariance. In the case of uncorrelated linear terms all measures would equal each other.

Measure		**Sum** ($\sum_{j=1}^{d} \cdot$)	**Value** ($\cdot$)	**Covariance**	**Description**
Variance partition	V_x	≥ 0	≥ 0	x	Proportion of variation due to the linear term
Diagonal variance partition	P_x	1	$[0, 1]$		Proportion of diagonal variance due to the linear term
Unique variance partition	U_x	$[0, 1]$	$[0, 1]$	x	Proportion of variation uniquely due to the linear term; square of the semi-partial correlation coefficient
Marginal variance partition	M_x	1	$[-d + 2; d]$	x	Marginal or averaged variance contribution of the linear term; equals the covariance between the linear term and the linear predictor.

Table 3. Joint interpretations of variance partitioning measures. V stands for the variance partition of a linear term or group of linear terms, M for its marginal variance partition and U for its unique variance partition. $\sum|r|$ is the sum of absolute values of the correlation coefficients between the linear term and all other linear terms and $\bar{r}$ is the mean of the those correlation coefficients.

Situation	Implication for other measures	Magnitude of correlation	Average correlation	Description	Interpretation
$V = 0$	$U = 0, M = 0$	0	0	There is no variation in the linear term.	The linear term is constant.
$V \approx 0, M \approx 0$	$U \approx 0$	$\sum\|r\| \approx 0$	$\bar{r} \approx 0$	The linear predictor, η, has no linear dependence on the linear term.	The linear term does not influence the variability of the linear predictor.
$V \approx U \neq 0$	$M \approx V$	$\sum\|r\| \approx 0$	$\bar{r} \approx 0$	The linear term is independent of the other linear terms.	The linear term has an independent effect on the variability of the linear predictor.
$M > V, V > 0$	$U < V$	$\sum\|r\| > 0$	$\bar{r} > 0$	Covariances between linear terms increase the total contribution of the linear term to the variance partitioning.	The linear term is confounded with other linear terms, such that only part of its variance contribution can be uniquely attributed to that linear term.
$M < V, V > 0$	$U < V$	$\sum\|r\| > 0$	$\bar{r} < 0$	If $M \approx 0$, the linear term has no direct effect on the variability of the linear predictor; if $M > 0$, the impact of the linear term is dampened by the covariances; if $M < 0$, the covariate acts as a suppressor on the linear predictor.	There is confounding and suppression between the linear terms. The overall effect on variability of the linear predictor depends on the sign of M.
$M \approx V, U < V$		$\sum\|r\| > 0$	$\bar{r} \approx 0$	The covariances average out.	The linear term is confounded with other linear terms, but the indirect effects cancel out.

Figures

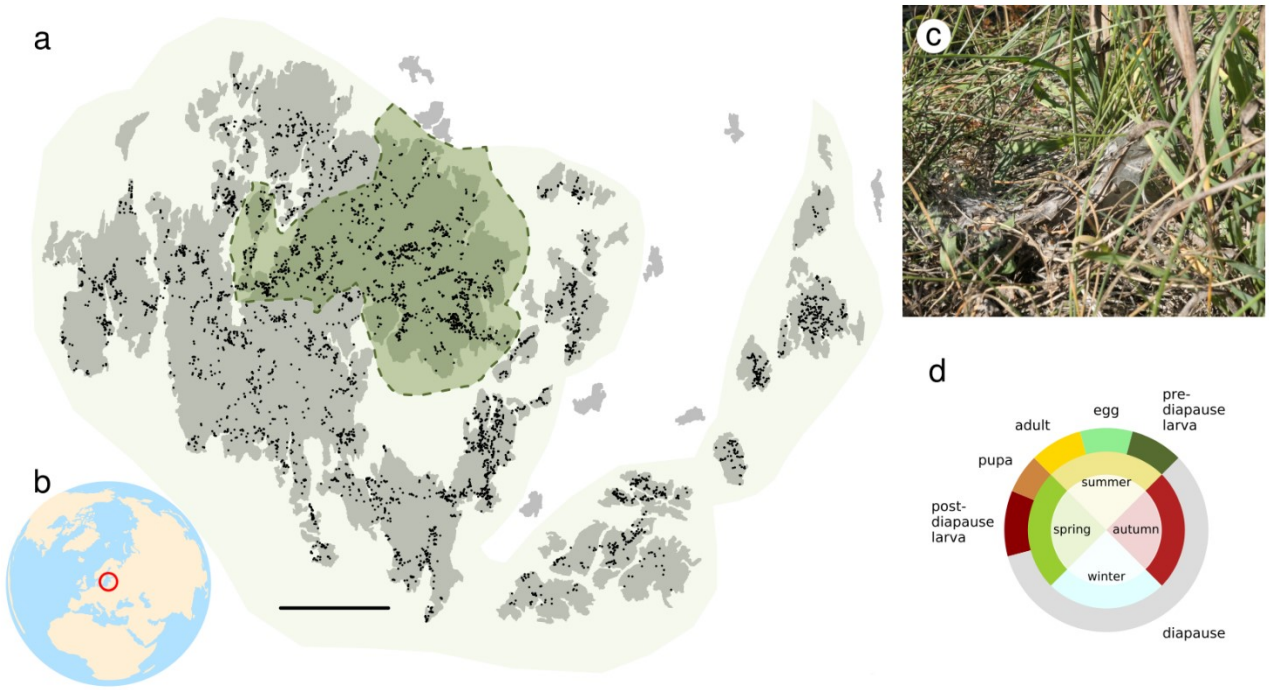

Figure 1. (**a**) Map of the Åland Islands study system. The black dots represent habitat patches suitable for *M. cinxia*. The darker enclosed region corresponds to the core area used for model inference, while the whole highlighted area corresponds to the "population" in our *population variance partitioning* measures. The solid black bar indicates 10 km in the map's projected coordinate system ETRS-TM35FIN. Shoreline data from GSHHG version 2.3.7. (**b**) A projected view of the globe with the Åland Islands at its center (lat. 58.7 N lon. 21.7 E). The red circle encloses approximate location of Åland in the Baltic Sea. Land area data from Natural Earth. (**c**) Silken 'winter nest' spun by the larvae of *M. cinxia* on *Plantago lanceolata*. Photograph by Torsti Schulz. (**d**) Life cycle of *M. cinxia*.

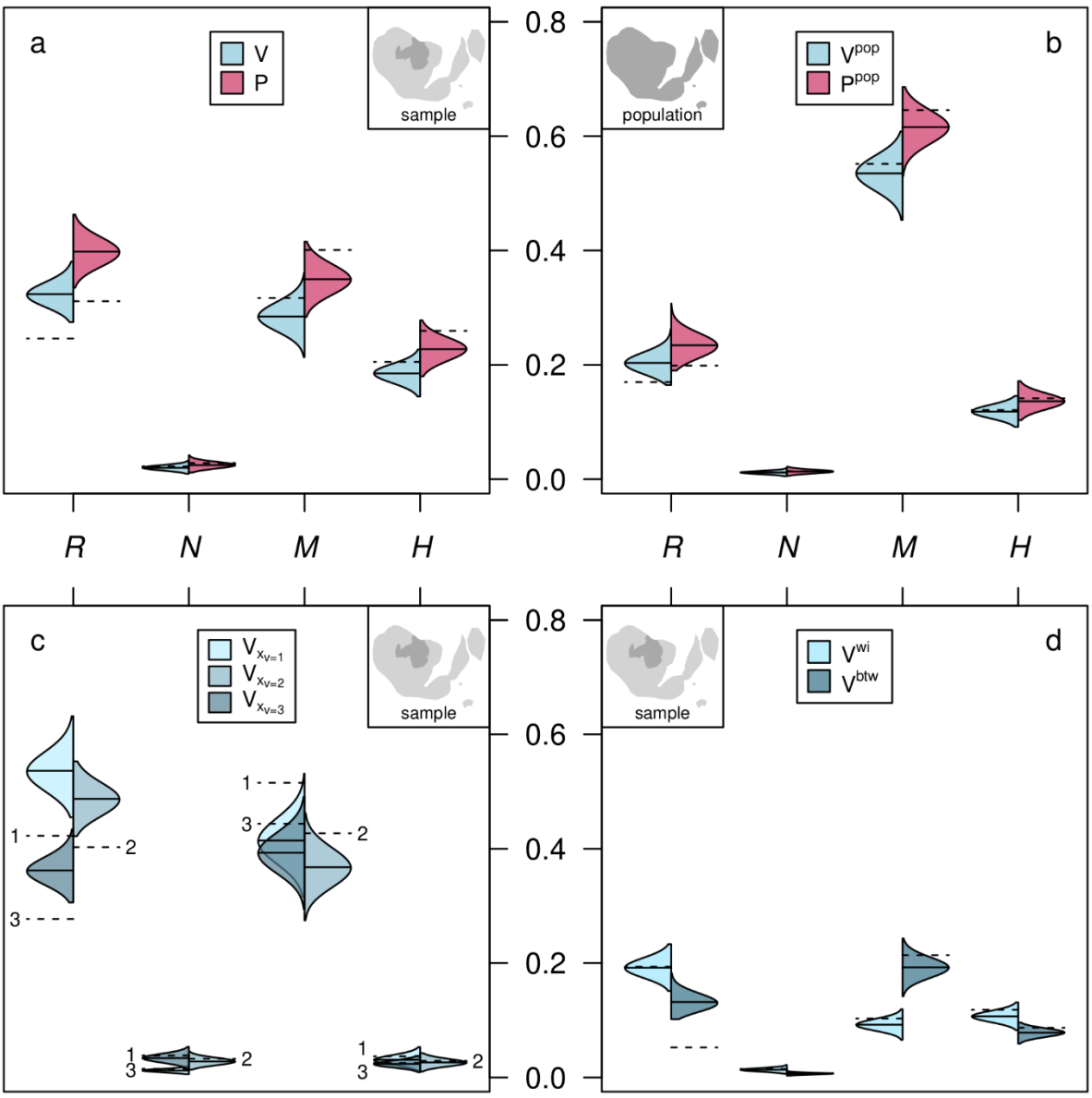

Figure 2. Variance partitioning of the linear predictor. The variance terms are grouped into four groups of linear terms: random effects (R) previous occupancy (N), metapopulation covariates (M), and habitat quality covariates (H). The violin plot distributions correspond to $p(V|D)$ and $p(P|D)$. Solid lines represent the posterior mean of $\hat{V}$ and $\hat{P}$ as the point estimate summary of the variance partition. Dashed lines represent the corresponding biased point estimate approximations with the posterior mean of the model parameters ($V(\hat{\theta})$ and $P(\hat{\theta})$). The inset maps show over which part of the study region the variance partition is calculated (see also Fig. 1). (**a**) Sample variance partitions and diagonal variance partitions for the core study area. (**b**) Variance partition including predictions to the survey areas not included in the model. (**c**) Sample variance partitioning conditional on the different levels of the host plant covariate (vegetation). (**d**) Sample variance partitions within and among patches.

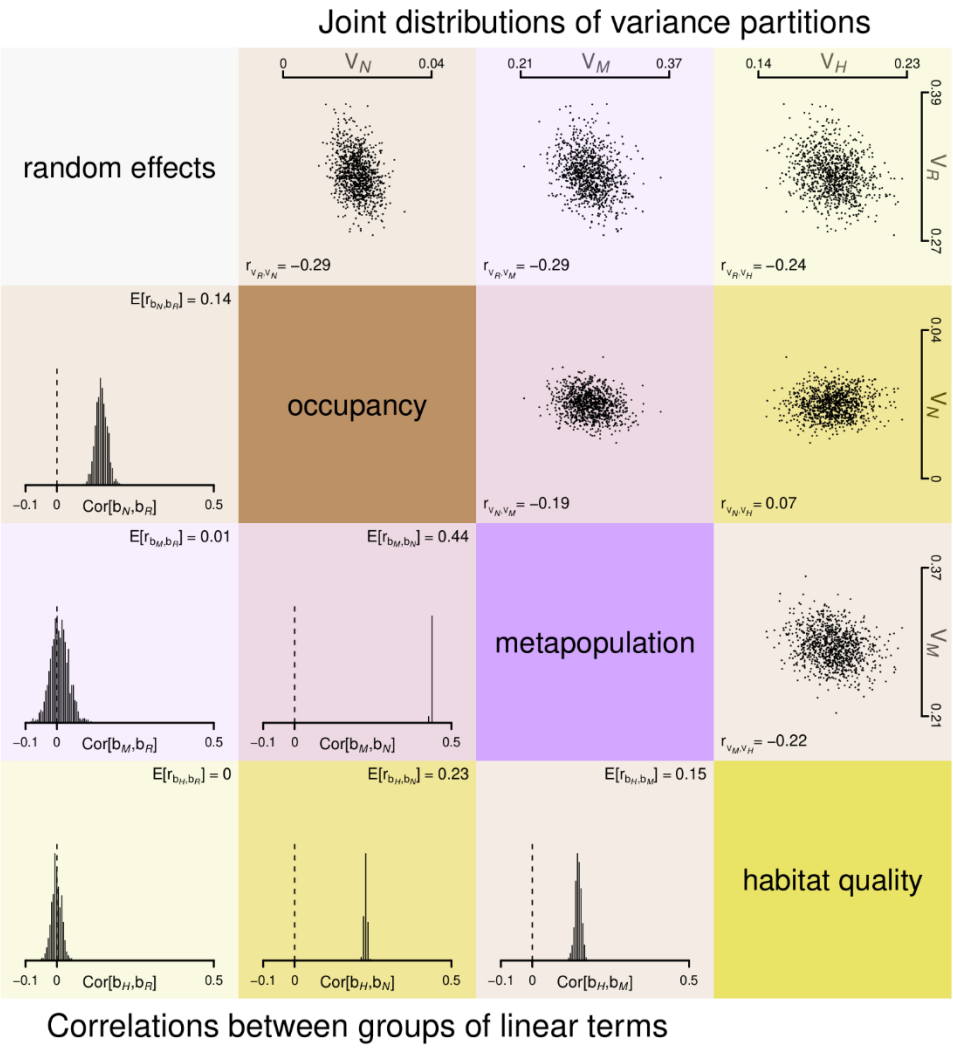

Figure 3. Posterior distributions of correlations between variance partitions and groups of linear terms for random effects (R) previous occupancy (N), metapopulation covariates (M), and habitat quality covariates (H). The histograms in the lower triangle represent the posterior marginal distributions of the correlations between the four groups of linear terms, $p(\mathrm{Cor}[b_l, b_{l'}]|D)$. Their posterior means are given in the upper right corner of each entry. The scatterplots in the upper triangle represent samples from the joint posterior distribution of pairs of variance partition measures ($p(V_{b_l}, V_{b_{l'}}|D)$). The posterior correlation between the variance partition measures is given in the lower left corner of each entry.

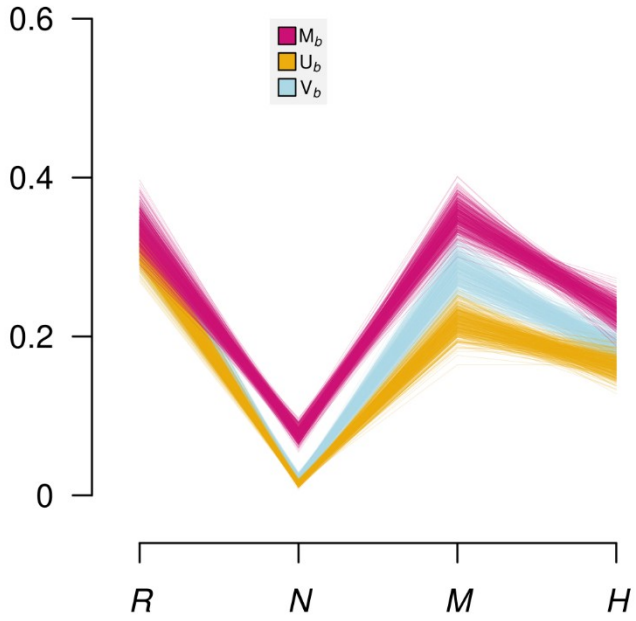

Figure 4. Parallel coordinates plot of the different measures of variance allocated to each group of linear terms. The groups are the random effects (R) previous occupancy (N), metapopulation covariates (M), and habitat quality covariates (H). The variance partitioning measures are the variance partition V_b in light blue, the unique variance partition U_b in orange and the marginal variance partition M_b in dark pink. Each line in the plot correspond to a sample from the (joint) posterior distribution of the measures.

Supplementary materials

Mathematical supplement

Programming supplement

Model supplement

Code supplement