Location-scale Meta-analysis and Meta-regression as a Tool to Capture Large-scale Changes in Biological and Methodological Heterogeneity: a Spotlight on Heteroscedasticity

Shinichi Nakagawa[1,2*], Ayumi Mizuno[1,2], Kyle Morrison[2], Lorenzo Ricolfi[2], Coralie Williams[2,3], Szymon M Drobniak[2,4], Malgorzata Lagisz[1,2], and Yefeng Yang[2*]

[1]Department of Biological Sciences, Faculty of Science, University of Alberta, Edmonton, Canada

[2]Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia

[3]School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia

[4]Institute of Environmental Sciences, Faculty of Biology, Jagiellonian University, Kraków, Poland

[*]shinichi.nakagawa@ualberta.ca & yefeng.yang1@unsw.edu.au

February 11, 2025

Abstract

Heterogeneity is a defining feature of ecological and evolutionary meta-analyses. While conventional metal-analysis and meta-regression methods acknowledge heterogeneity in effect sizes, they typically assume this heterogeneity is constant across studies and levels of moderators (i.e., homoscedasticity). This assumption could mask potentially informative patterns in the data. Here, we introduce and develop a location-scale meta-analysis and meta-regression framework that models both the mean (location) and variance (scale) of effect sizes. Such a framework explicitly accommodates heteroscedasticity (differences in variance), thereby revealing when and why heterogeneity itself changes. This capability, we argue, is crucial for understanding responses to global environmental change, where complex, context-dependent processes may shape both the average magnitude and the variability of biological responses. For example, differences in study design, measurement protocols, environmental factors, or even evolutionary history can lead to systematic shifts in variance. By incorporating hierarchical (multilevel) structures and phylogenetic relationships, location-scale models can disentangle the contributions from different levels to both location and scale parts. We further attempt to extend the concepts of relative heterogeneity and publication bias into the scale part of meta-regression. With these methodological advances, we can identify patterns and processes that remain obscured under the constant variance assumption, thereby enhancing the biological interpretability and practical relevance of meta-analytic results. Notably, almost all published ecological and evolutionary meta-analytic data can be re-analysed using our proposed analytic framework to gain new insights. Altogether, location-scale meta-analysis and meta-regression provide a rich and holistic lens through which to view and interpret the intricate tapestry woven with ecological and evolutionary data. The proposed approach, thus, ultimately leads to more informed and context-specific conclusions about environmental changes and their impacts.

Keywords— multilevel meta-analysis, phylogenetic meta-analysis, double-hierarchical model, generalized linear mixed-effects model, Bayesian statistics

1 Introduction

Meta-analysis has become an indispensable tool in ecology and evolutionary biology; it offers a means to synthesize results across diverse studies and to detect broad-scale patterns and biases (e.g., publication bias) that may be invisible at the level of individual investigations (Gurevitch et al., 2018; Nakagawa et al., 2017; Yang et al., 2022). Yet, the process of explaining heterogeneous datasets is fraught with challenges. Studies differ not only in their focal taxa, systems, and conditions but also in methodologies, measurement protocols, and analytical approaches. Such complexity leads to substantial heterogeneity in effect sizes, which could obscure underlying biological signals and hinder our understanding of global ecological change. Indeed, variation not due to sample size differences across studies frequently accounts for more than 90% (i.e., $I^2 > 0.9$) in ecological and evolutionary meta-analyses (Senior et al., 2016; but see Yang, Noble, et al., 2023; note that in medicine, $I^2 > 0.75$ is considered to be high; Higgins et al., 2003).

Conventional meta-analytic frameworks attempt to accommodate heterogeneity by introducing random effects and moderator variables. These approaches recognize that effect sizes are not identical and that moderators – such as climate gradients, habitat types, taxonomic groups, or methodological factors – may help explain some of the variance (Gurevitch et al., 2018; Nakagawa, Yang, et al., 2023; Nakagawa et al., 2017). However, linear models including standard meta-analysis and meta-regressions typically assume homoscedasticity, meaning that the variance of effect sizes remains constant across levels of these moderators (Viechtbauer & López-López, 2022). Such an assumption can be unrealistic, as both biological processes and methodological variation often influence not only the magnitude but also the variability of responses (Cleasby & Nakagawa, 2011). For example, under some environmental conditions, species or communities may display highly consistent responses, while in others, responses may be much more variable. Similarly, one type of measurement can be more consistent than another type of measurement.

In environmental sciences, including global change biology, this distinction between average responses and their variability is crucial. Understanding how variance patterns shift along environmental gradients or across study designs can illuminate processes of adaptation, resilience, or sensitivity (Pecl et al., 2017; Urban, 2015). For instance, certain anthropogenic changes, such as climate warming or habitat fragmentation, might not only alter the mean response of organisms but also produce more divergent responses among studies due to underlying differences in selection regimes, resource availability, or measurement uncertainty (Fig. 1; e.g., Pottier et al. 2022; Mathot et al. 2024). Without explicitly modelling the variance as a function of moderators, these subtle but important patterns of variability remain hidden (Cleasby & Nakagawa, 2011; Nakagawa et al., 2015; Senior et al., 2020).

Recent advances in statistical modelling offer solutions to this problem. Location-scale (mean-variance) modelling frameworks have been long recognized in other areas of statistics and quantitative genetics (Lee & Nelder, 1996, 2006; Rönnegård et al., 2010; Sae-Lim et al., 2015) and more recently, they have been adopted in ecology, evolution and environmental sciences (Cleasby et al., 2015; Mulder et al., 2016; O'Dea et al., 2022; Pitt et al., 2020). However, their application to meta-analysis, a domain inherently characterized by high-level heterogeneity, remains underexplored (Viechtbauer & López-López, 2022). We can directly model heteroscedasticity and partition the drivers of variability more explicitly, by extending the concept of meta-regression to include both location (mean) and scale parts (variance). Moreover, multilevel and phylogenetic extensions of location-scale models allow researchers to capture hierarchical structures and evolutionary histories that shape both the average effect sizes and their variation (dispersion).

In this paper, we present location-scale meta-analysis and meta-regression as a flexible, broadly applicable methodology for analysing ecological and evolutionary meta-analytic data (cf., Blowes 2024). We outline the theoretical foundation of the approach and illustrate how to incorporate moderators into both the mean and variance components. We then show how the framework can be adopted to accommodate multilevel and phylogenetic models. Additionally, we describe how the idea of heterogeneity in meta-analysis can be extended in the scale part and how regression-based methods can be expanded to test new types of publication biases in the scale part. We provide illustrative examples of model implementations to demonstrate the usefulness and insights that can be gained, with three different ecological datasets using statistical software, `R` (R Core Team 2024; for an online tutorial, see link). Finally, we discuss how our proposed methodology improves our understanding of global change biology and potentially better predicts the future impact of global changes by revealing patterns of variability that mirror complex ecological and evolutionary realities.

2 Theory

Below, we develop location-scale meta-analytic models of increasing complexity in five steps. These steps include extending the quantification of heterogeneity and the detection of publication bias from the mean part to the scale part (cf., Viechtbauer and López-López 2022).

2.1 Random-effects meta-analysis and meta-regression

The starting point of most ecological and evolutionary meta-analyses is the random-effects model (Nakagawa & Santos, 2012). Consider a set of studies indexed by $i = 1, \ldots, K$, each reporting an effect size y_i (i.e., one effect size per study). The random-effects model can be written as (Hedges, 1983):

$$y_i = \beta_0 + e_i + m_i, \tag{1}$$

where β_0 is the overall meta-analytic mean (intercept), e_i represents the study effect for i-th study (also, the i-th effect-size effect under this example, as the number of effect sizes and studies are the same), and m_i is the sampling error of the effect size estimate. Typically, we assume:

$$e_i \sim \mathcal{N}(0, \sigma_e^2), \tag{2}$$

and

$$m_i \sim \mathcal{N}(0, \sigma_{m_i}^2), \tag{3}$$

where σ_e^2 is the between-study (between-effect-size) variance, and $\sigma_{m_i}^2$ is sampling variance for i-th study (effect size) assumed to be known (often computed as a plug-in estimator from study-level sample sizes or other data). For example, when the effect size is a Fisher's z-transformed correlation coefficient z_r, the sampling variance often takes a simple form like $1/(n_i - 3)$, where n_i is the sample size of the i-th study.

Note that the random-effects model assumes different studies have different means (Hedges, 1983); if there is no (between-study) heterogeneity or $\sigma_e^2 = 0$, then the random-effects model reduces to the fixed-effect model where the overall mean (β_0) is the 'true' mean for all the studies. Also, note that σ_e^2 is hard to interpret as a general measure of heterogeneity because its magnitude depends on what type of effect size one uses. Therefore, the most common and relative measure of heterogeneity in meta-analysis (see Yang, Noble, et al. 2023) is:

$$I^2 = \frac{\sigma_e^2}{\sigma_e^2 + \overline{\sigma^2}_m} \tag{4}$$

with

$$\overline{\sigma^2}_m = \frac{\sum \frac{1}{\sigma_{m_i}^2} (K-1)}{\left(\sum \frac{1}{\sigma_{m_i}^2}\right)^2 - \sum \left(\frac{1}{\sigma_{m_i}^2}\right)^2}, \tag{5}$$

where $\overline{\sigma^2}_m$ is a typical (average) sampling variance (Higgins & Thompson, 2002; Higgins et al., 2003); note that to obtain I^2 or related indices, we use estimated parameters, i.e., variance components (e.g., via restricted maximum likelihood, REML, estimator or Bayesian estimators using Markov Chain Monte Carlo, MCMC). In this form,

(between-study) heterogeneity is expressed as a ratio in relation to the total variance (i.e., $\sigma_e^2 + \overline{\sigma^2}_m$; we extend the idea of heterogeneity in meta-analysis in Sections 2.3 and 2.4 below).

When moderators are introduced (e.g., $x_{1i}, x_{2i}, \ldots, x_{pi}$; i.e., having p moderators), the model extends to:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + e_i + m_i. \tag{6}$$

This standard meta-regression framework allows one to examine how moderators (covariates) influence the average effect size. However, it still maintains the assumption of a constant heterogeneity variance σ_e^2, ignoring potential differences in variance structure among different levels or values of the moderators. Note that a moderator (or predictor; x) can be a continuous or categorical variable.

We note that when a categorical moderator has a h levels, we have $h-1$ dummy variables (i.e., $h-1$ xs) and corresponding regression coefficients βs are usually contrasts (differences) between a reference category (level) and another category (level). A recent survey shows that almost all ecological and evolutionary meta-regression analyses had at least one categorical moderator (97%) while only around 30% of meta-regression analyses included at least one continuous moderator (Nakagawa, Lagisz, O'Dea, et al., 2023). This finding indicates dummy variables are very common in meta-regression analyses in ecology and evolution.

Given extremely high heterogeneities in ecological and evolutionary meta-analyses (Gurevitch et al., 2018), it is notable that meta-regression models, which include moderators (Equation 6), are the main analytical focus rather than meta-analytic models (i.e. intercept-only models; Equation 1). A significant moderator in a meta-regression is a piece of 'synthesis-generated evidence' because such evidence cannot be identified by examining each separate study (Cooper, 2015; Nakagawa et al., 2017) .

2.2 Random-effects location-scale meta-regression

Location-scale meta-regression explicitly models not only the location (mean) of the effects but also their scale (variance), allowing heteroscedasticity to be a function of moderators (cf., Cleasby and Nakagawa 2011). We extend the above meta-regression (Equation 6) by decomposing the model into a location part and a scale part (Viechtbauer & López-López, 2022):

$$y_i = \beta_0^{(l)} + \beta_1^{(l)} x_{1i} + \cdots + \beta_p^{(l)} x_{pi} + e_i^{(l)} + m_i^{(l)}, \tag{7}$$

with

$$e_i^{(l)} \sim \mathcal{N}(0, \sigma_{e_i}^2). \tag{8}$$

where $\beta_p^{(l)}$ are location parameters (i.e., affecting the mean part), and we allow the residual variance $\sigma_{e_i(l)}^2$ to vary by modelling the logarithm of its squared-root (i.e., standard deviation $\ln(\sigma_{e_i})$) as a linear function of moderators:

$$\ln(\sigma_{e_i}) = \beta_0^{(s)} + \beta_1^{(s)} x_{1i} + \cdots + \beta_p^{(s)} x_{pi}, \tag{9}$$

where $\beta_p^{(s)}$ coefficients indicate how much moderators influence heterogeneity itself.

In the scale part, any factor x_{pi} influencing the scale part ($\beta_p^{(s)} \neq 0$) implies that heterogeneity itself changes systematically with the moderator (i.e., heteroscedasticity). For example, a binary (categorical) moderator might lead to different levels having distinct variances (e.g., aquatic organisms having higher variance than terrestrial counterparts as in Example 1 below). Both the logarithm of the variance $\ln(\sigma_{e_i}^2)$ or standard deviation $\ln(\sigma_{e_i})$ can be the response variable in the scale part and the choice is a matter of preference; for example, O'Dea et al. (2022) uses

variance or $\ln(\sigma^2_{e_i})$ while Cleasby et al. (2015) uses standard deviation $\ln(\sigma_{e_i})$. We use $\ln(\sigma_{e_i})$ because our choice of implementation, the `R` package `brms` (Bürkner, 2017), uses standard deviation rather than variance. We should also note that a set of moderators does not need to be the same in the location and scale parts. Yet, without any clear prior predictions, one could start with the same moderators in both parts.

2.3 Multilevel meta-analysis and multilevel location-scale meta-regression

Many meta-analyses contain hierarchical structures, such as multiple effect sizes nested within studies (cf., Rodriguez et al. 2023; Williams et al. 2021). Indeed, a survey revealed that such a nested structure was present in 73 out of 73 meta-analytic studies (100%) in environmental sciences (Nakagawa, Yang, et al., 2023). Before introducing the multilevel location-scale meta-regression, we briefly review the standard multilevel meta-analytic model, which can be written as:

$$y_i = \beta_0 + u_{j[i]} + e_i + m_i, \tag{10}$$

with

$$u_j \sim \mathcal{N}(0, \sigma^2_u), \tag{11}$$

where $u_{j[i]}$ is the between-study effect for the j-th study (or of the i-th effect size) and σ_u is the between-study variance and e_i follows Equation 2, but it is notable that $\sigma^2_{e_i}$ is now the within-study variance (effect-size-level variance). Notably, m_i can be distributed following Equation 3, but it is more likely to take the following form:

$$m_i \sim \mathcal{N}(0, \mathbf{V}), \tag{12}$$

where $\mathbf{V}$ is a block diagonal matrix capturing the sampling covariance structure within and among effect sizes from the same study. For example, if we have 20 studies and, then, we have 20 blocks and, say, we can see the first 3 studies where they have 3, 1, 2 effect sizes, respectively. Let us further assume that the first 2 effect sizes in study 1 are derived from the same subjects and so are the two effect sizes in study 3 (elsewhere, we called such types of dependencies as correlated sampling errors to distinguish this dependence from another type of dependence due to belonging to the same studies, controlled by the random effect $u_{j[i]}$; see Nakagawa, Yang, et al. 2023; Yang, Macleod, et al. 2023). We can now write the first three blocks of $\mathbf{V}$ as (note that the boxes are drawn to show three blocks, which corresponds to three studies):

$$\mathbf{V_{1-3}} = \begin{pmatrix} \boxed{\begin{matrix} \sigma^2_{m_1} & \rho_m\sigma_{m_1}\sigma_{m_2} & 0 \\ \rho_m\sigma_{m_2}\sigma_{m_1} & \sigma^2_{m_2} & 0 \\ 0 & 0 & \sigma^2_{m_3} \end{matrix}} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ \boxed{\sigma^2_{m_4}} \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ \boxed{\begin{matrix} \sigma^2_{m_5} & \rho_m\sigma_{m_5}\sigma_{m_6} \\ \rho_m\sigma_{m_6}\sigma_{m_5} & \sigma^2_{m_6} \end{matrix}} \end{matrix} \end{pmatrix}. \tag{13}$$

where ρ_m is correlation between sampling variances (e.g., of effect size 1 and 2; $\sigma^2_{m_1}$ and $\sigma^2_{m_2}$); the value of ρ_m takes a value between 0 and 1 yet an exact value is unknown apart from some special conditions (e.g., effect size 1 and 2 shared a control group or we have access to original data so that we can sometimes obtain ρ_m directly; see Noble et al. 2017). Therefore, we often assume that either $\rho_m = 0.5$ or 0.8 (Noble et al., 2017; Pustejovsky & Tipton, 2022); note the variance-covariance matrix $\mathbf{V}$ can be easily constructed by, for example, the function `vcalc` in the `R` package `metafor` (Viechtbauer, 2010). Alternatively, robust variance estimators can be employed; this approach

offers flexibility in handling complex dependency structures among sampling errors, as we do not need to define the value of ρ_m. Interestingly, Pustejovsky and Tipton (2022) recommend the combined use of $\mathbf{V}$ and robust variance estimators (see also Hedges et al. 2010).

By extending the concept of the relative heterogeneity above (Equation 4), we can now define three types of I^2 (Nakagawa & Santos, 2012; Nakagawa, Yang, et al., 2023; Yang, Noble, et al., 2023):

$$I^2_B = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_e + \overline{\sigma^2}_m}, \tag{14}$$

$$I^2_W = \frac{\sigma^2_e}{\sigma^2_u + \sigma^2_e + \overline{\sigma^2}_m}, \tag{15}$$

and

$$I^2_T = \frac{\sigma^2_u + \sigma^2_e}{\sigma^2_u + \sigma^2_e + \overline{\sigma^2}_m}. \tag{16}$$

As one can see, I^2_B is the relative heterogeneity of between-study effects (differences), while I^2_W is that of within-study effects, and the sum of these two is I^2_T (total relative heterogeneity; for details and other types of relative heterogeneity measures, see Yang, Noble, et al. 2023).

Now we can define a multilevel location-scale meta-regression model building upon the multilevel meta-analytic model (Equation 10); the location part is:

$$y_i = \beta^{(l)}_0 + \beta^{(l)}_1 x_{1i} + \cdots + \beta^{(l)}_p x_{pi} + u^{(l)}_{j[i]} + e^{(l)}_i + m^{(l)}_i, \tag{17}$$

with

$$u^{(l)}_j \sim \mathcal{N}(0, \sigma^2_{u(l)}), \tag{18}$$

where $u^{(l)}_{j[i]}$ is the between-study effect for the j-th study that i-th effect size belongs to, $\sigma^2_{u(l)}$ is the between-study variance, and other symbols as above. The scale equation can be the same as Equation 9. Yet, it is important to notice that we could add the between-study effect to the scale part:

$$\ln(\sigma_{e_i}) = \beta^{(s)}_0 + \beta^{(s)}_1 x_{1i} + \cdots + \beta^{(s)}_p x_{pi} + u^{(s)}_{j[i]}. \tag{19}$$

By including the random effects (between-study effects) in both the location and scale equations and correlating them, we can model scenarios where studies with larger (or smaller) mean effects might also tend to exhibit greater (or smaller) variance; note that models with random effects in both location and scale parts are known as "double-hierarchical" models (Lee & Nelder, 1996, 2006). Formally, we can define a bivariate normal distribution for the between-study effects:

$$\begin{pmatrix} u^{(l)}_j \\ u^{(s)}_j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u(l)} & \rho_u \sigma_{u(l)} \sigma_{u(s)} \\ \rho_u \sigma_{u(l)} \sigma_{u(s)} & \sigma^2_{u(s)} \end{pmatrix} \right). \tag{20}$$

Here, the value of $\sigma^2_{u(s)}$ indicates the magnitude of differences in variance between studies (also a large value indicates a likely existence of heteroscedasticity; Fig. 2), and ρ_u measures the correlation between the location and scale

random effects, and unlike ρ_m in an earlier section, ρ_u spans between -1 and 1 (not between 0 and 1). Under normal circumstances, we do not expect any correlation between $u_j^{(l)}$ and $u_j^{(s)}$ because the default assumption is that mean and variance are independent in normally (Gaussian) distributed data (Fig. 2 showing different patterns of this correlation). Yet in biology, mean and variance may often be positively correlated, which is known as Taylor's law (Taylor 1961; see also Nakagawa et al. 2015). Of relevance, researchers have found that there is a positive correlation between sampling variance ($\sigma^2_{m_i}$) and heterogeneity (of means), equivalent measures of $\sigma^2_{u(s)}$ (Fig. 2f); that is, primary studies with smaller sample sizes tend to have larger heterogeneity (or larger residual value or $\ln(\sigma_{e_i})$ (IntHout et al., 2015; Stanley et al., 2022). Given small studies often have large effect sizes in magnitudes, this finding indicates that we may find that larger effects in magnitude are related to high variance in a meta-analysis (i.e., non-zero ρ_u between effect sizes and heterogeneity), a pattern that may suggest selection bias or other methodological artefacts (e.g., smaller studies reporting both inflated means and noisier/variable results; Stanley et al. 2022). Notably, larger studies are less likely to be affected by these issues, and thus, large-study divergence is unlikely to occur as mentioned earlier.

Notably, adding the between-study (random) effect in the scale part results in two extra parameters to estimate, i.e., $\sigma^2_{u(s)}$ and ρ_u; in addition to j between-study effects $u_i^{(s)}$, which naturally requires more data. Therefore, such a location-scale meta-regression model with the between-study effect in the scale part may require larger meta-analytic datasets (for more discussion, see our examples below). Nevertheless, it could be informative to estimate $\sigma^2_{u(s)}$ and ρ_u regardless of dataset size. Therefore, we suggest before fitting a multilevel location-scale meta-regression, we can first fit the following meta-analytic model:

$$y_i = \beta_0^{(l)} + u_{j[i]}^{(l)} + e_i^{(l)} + m_i^{(l)}, \tag{21}$$

and

$$\ln(\sigma_{e_i}) = \beta_0^{(s)} + u_{j[i]}^{(s)}. \tag{22}$$

This meta-analysis provides a more accurate error estimate of the overall effect (i.e., β_0) when there exists non-negligible variation in variance. We propose that this meta-analytic model should be the starting point if one is to investigate heteroscedasticity. This is because non-zero $\sigma^2_{u(s)}$ warrants location-scale meta-regression in the same way as heterogeneity in a normal meta-analytic model calls for a (standard) meta-regression analysis (Nakagawa & Santos, 2012).

Additionally, in this location-scale meta-analytic model (Equations 21-22), both $\sigma^2_{u(s)}$ and ρ_u can be estimated as in Equation 20. Yet, in location-scale models with the between-study effects in both parts (i.e., double-hierarchical models), it is possible not to model ρ_u by assuming $\rho_u = 0$ as below, especially, when modelling ρ_u leads to difficulties in model convergence (which could help convergence and mixing in a Bayesian model; see the examples below and the online tutorial):

$$\begin{pmatrix} u_j^{(l)} \\ u_j^{(s)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u(l)} & 0 \\ 0 & \sigma^2_{u(s)} \end{pmatrix} \right). \tag{23}$$

It is interesting and maybe insightful to compare these two kinds of heterogeneity: 1) heterogeneity in mean and 2) heterogeneity in variance. Yet, we cannot compare these two parameters directly because they are on different scales (i.e., the latter is on the log-normal scale). In the next section, we resolve this very issue.

2.4 Extending the idea of heterogeneity to location-scale models

Earlier, we introduced the relative measure of heterogeneity, I^2 (variance-standardized measure). For Equation 21 (the location part), we can also calculate three types of I^2 as with Equations 14-16. For example, I^2_B can be obtained as:

$$I^2_B = \frac{\sigma^2_{u(l)}}{\sigma^2_{u(l)} + \overline{\sigma^2}_e + \overline{\sigma^2}_m}, \tag{24}$$

with

$$\overline{\sigma^2}_e = \exp\left(2\beta_0^{(s)} + 2\sigma^2_{u(s)}\right) \tag{25}$$

where $\beta_0^{(s)}$ is from Equation 22 and $\sigma^2_{u(s)}$ is the variance component for the between-study effects $u^{(s)}_{j[i]}$ from the same equation (O'Dea et al., 2022). Yet, it is not possible to extend I^2 to the scale part as the part lacks an equivalent of the sampling error variance (i.e., $\sigma^2_{m_i}$). Although less used, there is an alternative measurement of relative heterogeneity for meta-analysis, which is mean-standardized (Cairns & Prendergast, 2022). Using the random-effects model (Equation 1), we can define this measure (CV_H) as (Takkouche et al., 1999):

$$CV_H = \frac{\sigma_e}{|\beta_0|}, \tag{26}$$

where CV denotes the coefficient of variation, and $|\beta_0|$ is the absolute values of the overall mean (they match $|\beta_0|$ in Equation 1 but not necessarily in meta-regression models).

For a multi-level meta-analysis (Equation 10), we have (Yang, Noble, et al., 2023):

$$CV_{H(B)} = \frac{\sigma_u}{|\beta_0|}, \tag{27}$$

$$CV_{H(W)} = \frac{\sigma_e}{|\beta_0|}, \tag{28}$$

and

$$CV_{H(T)} = \frac{\sqrt{\sigma^2_e + \sigma^2_u}}{|\beta_0|}, \tag{29}$$

where $CV_{H(B)}$, $CV_{H(W)}$, and $CV_{H(T)}$ are between-study, within-study and total relative heterogeneity although $CV_{H(B)} + CV_{H(W)} \neq CV_{H(T)}$ (but $CV^2_{H(B)} + CV^2_{H(W)} = CV^2_{H(T)}$; Yang, Noble, et al. 2023; cf., Equations 14-16).

Mentioned earlier, the location-scale meta-analytic model in the previous section (Equations 21-22) has the between-study effects in both the location and scale part. We can, therefore, define relative heterogeneity (CV_H) for both the location and scale parts, using Equations 21-22:

$$CV^{(l)}_{H(B)} = \frac{\sigma_{u(l)}}{|\beta_0|}, \tag{30}$$

and

$$CV^{(s)}_{H(B)} = \sqrt{\exp(\sigma^2_{u(s)}) - 1}, \tag{31}$$

where $CV^{(l)}_{H(B)}$ and $CV^{(s)}_{H(B)}$ are between-study relative heterogeneity for the location and scale part, respectively. Although Equation 31 does not look like a coefficient of variation, it indeed is (see Cleasby et al. 2015; O'Dea et al.

2022). These two types of CV can be comparable in theory (yet note that these measures were originally developed for ratio scale variables, which have zero as the minimum value). For example, if both CVs are similar values, variability in mean and heterogeneity are similar (see Fig. 2). We note that these measures have yet to be used in meta-analyses, so it is hard to gauge their usefulness (cf., Yang, Noble, et al. 2023). Yet, the consistency of studies in terms of mean and variance should be of importance for many meta-analysts.

2.5 Modelling four types of publication bias in location-scale models

Publication biases, such as small-study effects and decline effects, can influence meta-analytic results (Rothstein et al., 2005). The small-study effect happens when selective publications of small studies with only significant effects, biasing an overall mean. The decline effect occurs when larger and statistically significant effects are published earlier than smaller and non-statistically significant effects, resulting in a decline in the magnitude of the overall effect over time (also known as time-lag bias; Koricheva and Kulinskaya 2019); while an incline effect may theoretically possible, practically, it is rarely, if ever, observed (Yang, Lagisz, & Nakagawa, 2023). Indeed, both types of publication bias are common in ecology and evolution (Yang, Sánchez-Tójar, et al., 2023; Yang et al., 2022). One of the notable strengths of meta-analysis is its ability to detect such publication biases.

For example, small-study effects can be examined by regressing y_i on the square root of sampling variance (standard error, se; Egger et al. 1997; Moreno et al. 2009).

$$y_i = \beta_0 + \beta_1 se_i + \cdots + u_{j[i]} + e_i + m_i, \tag{32}$$

where se_i is sampling standard deviation for i-th effect size (the square root of sampling variance, also often referred to as sampling standard error; for Zr, it $1/(n_i - 3)$). Alternatively, we can use $\sqrt{1/\tilde{n}_i}$, where $\tilde{n}_i$ is an effective sample size for i-th effect size and the use of such effective sample size avoids known correlation between effect size point estimates and their standard error as in standardized mean difference, SMD (more often referred to as Cohen's d or Hedges' g; see Nakagawa et al. 2022):

$$y_i = \beta_0 + \beta_1 \sqrt{1/\tilde{n}_i} + \cdots + u_{j[i]} + e_i + m_i. \tag{33}$$

Without the presence of a small-study effect (publication bias), there should be no relationship between effect sizes and se_i (or $\sqrt{1/\tilde{n}_i}$), which form a funnel shape by effect size values converging to an overall value as se_i (or $\sqrt{1/\tilde{n}_i}$) increases. If $\beta_1 \neq 0$, this suggests funnel asymmetry and hence the small-study effect. A funnel asymmetry could happen due to other moderators than the effective sample size. Therefore, it is important to model other moderators, which account for variation in the data.

Similarly, the decline effect can be examined by including a centred publication year $c(\text{year}_i)$ as a moderator (note that centring is not essential yet helps interpretation; see Schielzeth 2010):

$$y_i = \beta_0 + \beta_1 c(\text{year}_i) + u_{j[i]} + \cdots + e_i + m_i. \tag{34}$$

By combining these moderators ($se_i/\sqrt{1/\tilde{n}_i}$ and $c(\text{year}_i)$), we can model both location and scale to detect how biases affect not only average effect sizes but also their heterogeneity. The location-scale version might look like (cf., Viechtbauer and López-López 2022):

$$y_i = \beta_0^{(l)} + \beta_1^{(l)} \sqrt{1/\tilde{n}_i} + \beta_2^{(l)} c(\text{year}_i) + \cdots + u_{j[i]}^{(l)} + e_i^{(l)} + m_i^{(l)}, \tag{35}$$

and

$$\ln(\sigma_{e_i}) = \beta_0^{(s)} + \beta_1^{(s)} \sqrt{1/\tilde{n}_i} + \beta_2^{(s)} c(\text{year}_i) + \cdots . \tag{36}$$

If $\beta_1^{(s)}$ is statistically significant, it implies that heterogeneity increases with decreasing sample size (often linked to small-study effects; IntHout et al. 2015; Viechtbauer and López-López 2022), whereas a significant $\beta_2^{(s)}$ might indicate a "Proteus" effect, where variance (heterogeneity) in effect sizes decline over time (Trikalinos & Ioannidis, 2005). The reason for the Proteus effect is that initially, it is easier to publish papers that contradict the initial findings, which leads to high variance initially. Still, over time, variance in effect sizes declines as a consensus emerges (Trikalinos & Ioannidis, 2005). However, in ecology and evolution, we predict that heterogeneity can increase over time because an initial finding in one population (or one species) is often tested in more populations (and more species), increasing variability in effect sizes over time. This is the opposite of what the original Proteus effect meant, expanding what a Proteus effect means to any changes in effect sizes over time.

Therefore, using Equation 35-36, we can quantify: a) a small-study effect (the location part; Fig. 3a), b) a decline effect (the location part; Fig. 3b), c) a small-study effect on variance, which we name 'small-study divergence' (it could be 'small-study' convergence, but it is unlikely see below; the scale part; Fig. 3c), and d) a Proteus effect (the scale part; Fig. 3d). Such comprehensive examinations have not been tried but can be valuable for diagnosing publication biases in meta-analytic data.

2.6 Phylogenetic (multilevel) location-scale meta-analysis and meta-regression

Ecological and evolutionary meta-analyses often deal with species-level data, where evolutionary history can shape both the mean and variance of effect sizes (Cinar et al., 2022; Hadfield & Nakagawa, 2010; Nakagawa & Santos, 2012). By building upon the multilevel model (Equation 10), a phylogenetic multilevel meta-analytic model can be written as:

$$y_i = \beta_0 + a_{k[i]} + s_{k[i]} + u_{j[i]} + e_i + m_i \tag{37}$$

with

$$a_k^{(l)} \sim \mathcal{N}(0, \sigma^2_{a(l)}\mathbf{A}), \tag{38}$$

and

$$s_k^{(l)} \sim \mathcal{N}(0, \sigma^2_{s(l)}). \tag{39}$$

where $a_{k[i]}^{(l)}$ captures the phylogenetic effect for the k-th species, and $s_{k[i]}^{(l)}$ is the non-phylogenetic (species-level random) effect for the k-th species, each of them is normally distributed with $\sigma^2_{a(l)}\mathbf{A}$ and $\sigma^2_{s(l)}$ and $\mathbf{A}$ is a correlation matrix containing relatedness of k species (Cinar et al., 2022; Hadfield & Nakagawa, 2010; Nakagawa & Santos, 2012). It is notable that the ratio between $\sigma^2_{a(l)}$ and $\sigma^2_{s(l)}$ can quantify the relative strength of 'phylogenetic signal' in a dataset. It is known either as λ or phylogenetic heritability (H^2; Lynch 1991;Cinar et al. 2022; but see Pearse et al. 2023;):

$$\lambda = H^2 = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_s}. \tag{40}$$

Based on the above, a phylogenetic location-scale meta-regression model can be written as:

$$y_i = \beta_0^{(l)} + \beta_1^{(l)} x_{1i} + \cdots + \beta_p^{(l)} x_{pi} + a_{k[i]}^{(l)} + s_{k[i]}^{(l)} + u_{j[i]}^{(l)} + e_i^{(l)} + m_i^{(l)}, \tag{41}$$

and the scale component can similarly incorporate moderators:

$$\ln(\sigma_{e_i}) = \beta_0^{(s)} + \beta_1^{(s)} x_{1i} + \cdots + \beta_p^{(s)} x_{pi} + \cdots . \tag{42}$$

By doing so, we can test whether certain clades or evolutionary lineages exhibit inherently different levels of heterogeneity by species-level moderators (e.g., two different species groups according to their taxonomy). This phylogenetic extension helps to unravel how evolutionary history, along with environmental or methodological moderators, shapes both the magnitude and dispersion of ecological and evolutionary responses. Note that we could add the between-study effect ($u_i^{(s)}$) in the scale part. Also, it is possible even to incorporate the phylogenetic and non-phylogenetic effects ($a_{k[i]}^{(s)}$ and $s_{k[i]}^{(s)}$) in the scale part, but we would not recommend such models unless one has a relatively large dataset; the more complex the model, the more data points are required (cf., Cinar et al. 2022). Also, the location part of a meta-analytic model can be written as:

$$y_i = \beta_0^{(l)} + a_{k[i]}^{(l)} + s_{k[i]}^{(l)} + u_i^{(l)} + e_i^{(l)} + m_i^{(l)}, \tag{43}$$

with the scale part being Equation 22; such a phylogenetic multilevel location-scale meta-analytic model can be run before fitting a meta-regression counterpart.

3 Worked Examples

Here, we provide illustrative examples by re-analysing data from three published meta-analyses. Our aim here is to show examples of models we described above, and, therefore, we note that our model structure (e.g., the absence of phylogenetic relatedness) and the choice of moderators are unlikely to be biologically or methodologically the best given these datasets. That is, our examples may present models that could be too simplistic and fail to fully capture the complexities of these datasets. For implementation, we primarily use the R package, `brms` (Bürkner, 2017) but for some models, we also use `metafor` (Viechtbauer, 2010) and `blsmeta` (Rodriguez et al., 2023; Williams et al., 2021); note that results from all three packages `brms`, `metafor` and `blsmeta` are all consistent to each other. The full `R` scripts, along with the datasets, are available on our tutorial page (link), which can also serve as an introduction to fitting standard meta-analysis and meta-regression using these packages. Notably, in the online tutorial, we start each example fitting a multilevel location-scale meta-analytic model, which we have recommended, above, as a starting point of modelling (i.e., Equations 21 and 22). Below,however, we focus on results from location-scale meta-regression models, mainly using Equations 17 and 9 rather than Equations 17 and 19; this is because the former mix and converge more easily and also multiple R packages can fit this model, although the latter can be a better model in some cases (note that it is possible to decide which model is better using Bayesian model selection using, for example, Widely Applicable Information Criterion, WAIC or leave-one-out cross-validation, loo-cv; Vehtari et al. 2017; see also Blowes 2024).

3.1 Example 1: Biological and Methodological Categorical Moderators

Pottier et al. (2022) studied the capacity of animals to increase thermal tolerance via heat exposure (increased temperature) using a meta-analysis with the ratio of acclimation response between control and heat-exposed groups, as effect sizes. Using multilevel (i.e., Equations 17 and 9), we re-analysed their dataset, whether `habitat` (living aquatic [aqu.] vs. terrestrial [ter.] habitat) and 'method' (experiments testing either early/initial exposure [ini.] or persistent exposure [per.]) moderate not only the mean effect but also variances. Indeed, not only terrestrial organisms had significantly lower heat tolerance overall than aquatic counterparts, overall ($\beta_{[\text{ter.-aqu.}]}^{(l)}$: -0.16, 95% CI: -0.23 to -0.29) but also they had significantly lower variability ($\beta_{[\text{ter.-aqu.}]}^{(s)}$: -1.18, 95% CI: -1.33 to -1.02; Fig. 4a). Also, persistent exposures, overall, increased heat tolerance yet, significantly less than early (initial) exposure ($\beta_{[\text{per.-ini.}]}^{(l)}$: -0.07, 95% CI: -0.10 to -0.03), although persistent exposures generated significantly more variability ($\beta_{[\text{per.-ini.}]}^{(s)}$: 0.21, 95% CI: 0.07 to 0.34; Fig. 4b). These reanalyses highlight the often neglected roles of biological and methodological moderators in meta-analyses; we expect and predict heteroscedasticity (i.e., significant contrasts (slopes) on the scale part ($\beta^{(s)}$) are prevalent in ecological and evolutionary meta-analytic data.

3.2 Example 2: A Continuous Biological Moderator

Midolo et al. (2019) examined how plant traits change along a relevant gradient, using log response ratio (lnRR; comparing trait differences over differences in `elevatoin`). Here, we re-analysed one of the traits, nitrogen concentration per unit of area (N*area*), using a location-scale meta-regression model. As with the original authors, we found an increase in elevation difference accompanied by a significant increase in N*area* ($\beta^{(l)}_{[\text{elevation}]}$: 0.05, 95% CI: 0.01 to 0.08)). More importantly, variances among effect sizes (lnRR for N*area*) also increased as the elevation increased ($\beta^{(s)}_{[\text{elevation}]}$: 0.29, 95% CI: 0.12 to 0.46); Fig. 4c). Although continuous moderators like elevation here are less common in meta-analytic data sets (Nakagawa, Lagisz, O'Dea, et al., 2023), heteroscedasticity in such moderators may be more common than we assume (i.e., homoscedasticity).

3.3 Example 3: Modelling Publication Bias in the Location and Scale Part

Neuschulz et al. (2016) studied the effect of forest disturbance on pollination, seed dispersal, seed predation, recruitment and herbivory during plant regeneration, using a meta-analysis with standardised means difference (SMD) as their effect size. We used their data set to test the four types of publication biases described above by fitting sampling standard error (`se`; note the higher the standard error, usually, the smaller the sample size) and the centred publication year (`cyear`). Although we did not find little statistical evidence for the small-study effect and the decline effect ($\beta^{(l)}_{[\text{se}]}$: -0.89, 95% CI: -2.06 to 0.23); $\beta^{(l)}_{[\text{cyear}]}$: -0.04, 95% CI: -0.12 to 0.04)), we found such evidence for small-study divergence ($\beta^{(s)}_{[\text{se}]}$: -0.19, 95% CI: 0.31 to -0.09) as well as the Proteus effect with variance going down over time ($\beta^{(s)}_{[\text{cyear}]}$: 2.13, 95% CI: 0.74 to 3.41; Fig. 4d). This example points out that the current practice of just testing for the small study and the decline effect may miss the complexity of publication bias, missing the important insights gained by testing publication bias on the scale part, i.e., the small-study divergence and Proteus effect.

4 Discussion

In this paper, we have introduced (phylogenetic) multilevel location-scale meta-analysis and meta-regression as a new methodological advance to better capture, understand, and interpret heterogeneity and heteroscedasticity in ecological and evolutionary meta-analyses with illustrative examples from global change biology (cf., Viechtbauer and López-López 2022; Blowes 2024). By jointly modelling the location (mean) and scale (variance) of effect sizes, this approach surpasses conventional frameworks that treat variance as a single, homogeneous quantity. Below, we highlight the key advantages and implications of this framework in eight points.

First, the location-scale framework enhances biological interpretability. Variability in responses is not merely noise; it can reflect underlying ecological and evolutionary processes. When variance differs systematically across moderators, we understand whether certain environments, taxa, or conditions channel responses into restricted or variable outcomes. Such insights are highly relevant in a rapidly changing world, where both shifting averages and expanding or contracting variances across populations may signal adaptive capacity, vulnerability, or underlying ecological complexity (Pecl et al., 2017; Urban, 2015). Notably, such changes in variation in response can be easily visualised by orchard plots (for categorical variables) or bubble plots (for continuous variables; Nakagawa et al. 2021, 2023; see Fig. 4).

Second, location-scale modelling helps disentangle methodological sources of heterogeneity. Differences in study design, measurement techniques, or analytical choices may inflate the variance of reported effect sizes (cf., Dougherty and Shuker 2015; Christie et al. 2019; Mathot et al. 2024). Incorporating methodological moderators into the scale component allows us to identify when and how systematic sources of variability arise, guiding future research toward more consistent protocols and improving the overall reliability and comparability of meta-analytic findings (Blowes, 2024).

Third, related to the first two points, we can also inform predictions, for example, under global change scenarios. As environmental drivers intensify, understanding not just how mean responses shift but also how variance itself changes is critical. Increased variability may indicate an ecological opportunity for some species or impending instability

for others. Modelling changes in variance gives us an additional tool to anticipate the directions and magnitudes of uncertainty that will accompany shifts in mean responses, ultimately improving our ability to forecast and manage biological responses to global change.

Fourth, integrating hierarchical (multilevel) structures into location-scale models accommodates ecological and evolutionary meta-analytic datasets with multiple effect sizes per study (cf., Viechtbauer and López-López 2022). This approach not only provides a clearer picture of the relative contributions of study-level and effect-level factors but also elucidates between-study heterogeneity in the scale part as well as in the location part (Yang, Noble, et al., 2023). Indeed, we have proposed a multilevel location-scale meta-analytic model with the between-study effects in both parts as the starting point for exploring heterogeneity in mean and variance (e.g., comparing $CV^{(l)}_{H(B)}$ and $CV^{(s)}_{H(B)}$).

Fifth, incorporating phylogenetic structures into location-scale models not only controls for nuisance non-independence but also deepens our evolutionary understanding (Cinar et al., 2022; Hadfield & Nakagawa, 2010; Nakagawa & Santos, 2012). By accounting for shared ancestry, we can determine whether specific clades inherently produce more variable responses, possibly due to broader genetic diversity, greater plasticity, or more complex ecological interactions. Phylogenetic extensions allow us to identify evolutionary patterns in both mean effect sizes and their variability.

Sixth, the location-scale framework enables more comprehensive investigations of publication biases; we have outlined the four types of publication biases (the small-study effect, decline effect, small-study divergence, and Proteus effect). Traditional tests focus on detecting biases in mean effect sizes (Koricheva and Kulinskaya 2019; Nakagawa et al. 2022). By including moderators in the scale component, we can also examine biases in heterogeneity itself. For instance, we may identify when small studies or more recent publications not only inflate mean effects but also increase variance, revealing previously undetected dimensions of bias. Such multifaceted examinations of publication biases can improve the robustness and trustworthiness of meta-analytical conclusions.

Seventh, therefore, the multifaceted approach enhances the interpretability of meta-analytic findings for stakeholders and policymakers (Koricheva & Kulinskaya, 2019; Yang, Noble, et al., 2023). Rather than presenting a single mean effect size with a uniform measure of heterogeneity, we can specify when and where heterogeneity increases or decreases. These more detailed insights can guide resource allocation, monitoring efforts, and mitigation strategies for conditions associated with the greatest uncertainties or susceptibilities.

Eighth, more broadly, location-scale meta-analytic models present an opportunity for synthesis and comparability across a wide range of ecological and evolutionary contexts. By applying this method to various research questions, we can begin to build a general understanding of how heterogeneity responds to both biological and methodological factors (cf., Cleasby and Nakagawa 2011). This holistic approach promises to enrich our grasp of biodiversity, ecosystem functioning, and evolutionary potential as they unfold under changing environmental conditions. Importantly, given reasonable sample sizes (e.g., 40 effect sizes; indicated by simulation in Rodriguez et al. 2023), all published ecological and evolutionary meta-analyses can be re-analysed with our proposed models to investigate heteroscedasticity.

In summary, location-scale meta-analysis and meta-regression models, with multilevel, phylogenetic, and publication-bias extensions, provide a versatile and biologically interpretable framework for meta-analysis. They allow researchers to understand how moderators influence average effect sizes and reveal the conditions under which heterogeneity is amplified or diminished. This yields deeper ecological and evolutionary insights, refines our interpretations of meta-analytic results, and ultimately advances our understanding of complex biological responses to global environmental change.

5 Acknowledgments

SN and AM were supported by a Canada Excellence Research Chair (CERC-2022-00074). Also, YY, SN, and ML were supported by a Australian Research Council Discovery Grant (DP230101248).

6 Figure

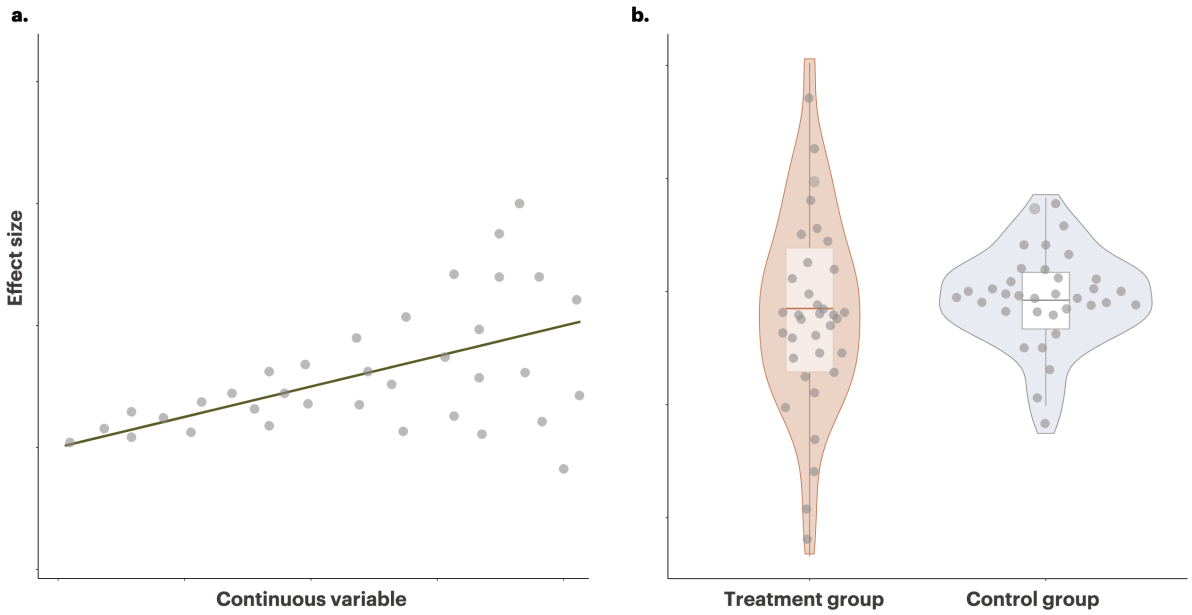

Figure 1: Visualizing heteroscedasticity: (a) an example of a continuous moderator (e.g., temperature, elevation, or sampling effort) with variance in effect sizes increase as moderator values increase. (b) an example of a categorical moderator (e.g., treatment vs. control groups or females vs. males) with the treatment group having more variation in effect sizes than the control group.

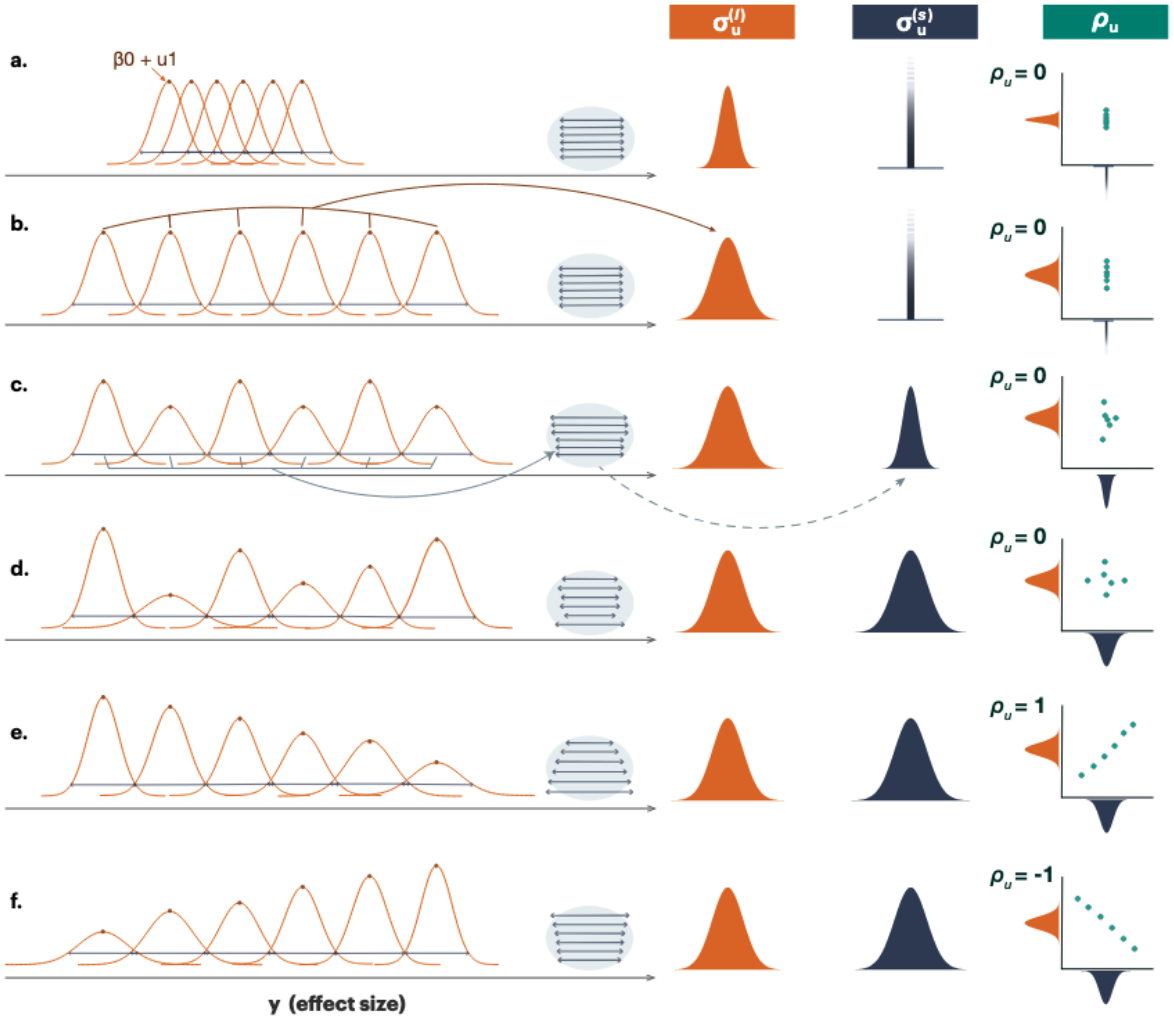

Figure 2: Illustration of location-scale models with different combinations of random effects in the location and scale parts. (a) Depicts a scenario where each study (orange curves) has its mean (i.e., $\beta_0 + u_j^{(l)}$) and variance (i.e., $u_j^{(s)}$). Between-study variation in the average effect size is represented by $\sigma^2_{u(l)}$ (light-orange distribution), and between-study variation in variance is represented by $\sigma^2_{u(s)}$ (navy distribution), which is zero (no variation or homoscedasticity). Correlation between these two random effects (ρ_u) can be zero, positive, or negative, leading to different patterns (in this case, zero), (b) in this scenario, everything is the same apart from between-study variance in means are larger than scenario a, (c) in this case, there is variations in variation, (d) in this case, each study differs in mean and variance with $\rho_u = 0$. (e) a positive correlation, $\rho_u = 1$, means that higher mean effects co-occur with greater variance. (f) a negative correlation, $\rho_u = -1$, means that higher mean effects co-occur with lower variance. Each panel on the right shows a schematic distribution of $u_j^{(l)}$ (orange) and $u_j^{(s)}$ (blue), along with their correlation in a scatterplot. These scenarios highlight how location-scale approaches can capture diverse patterns of heterogeneity.

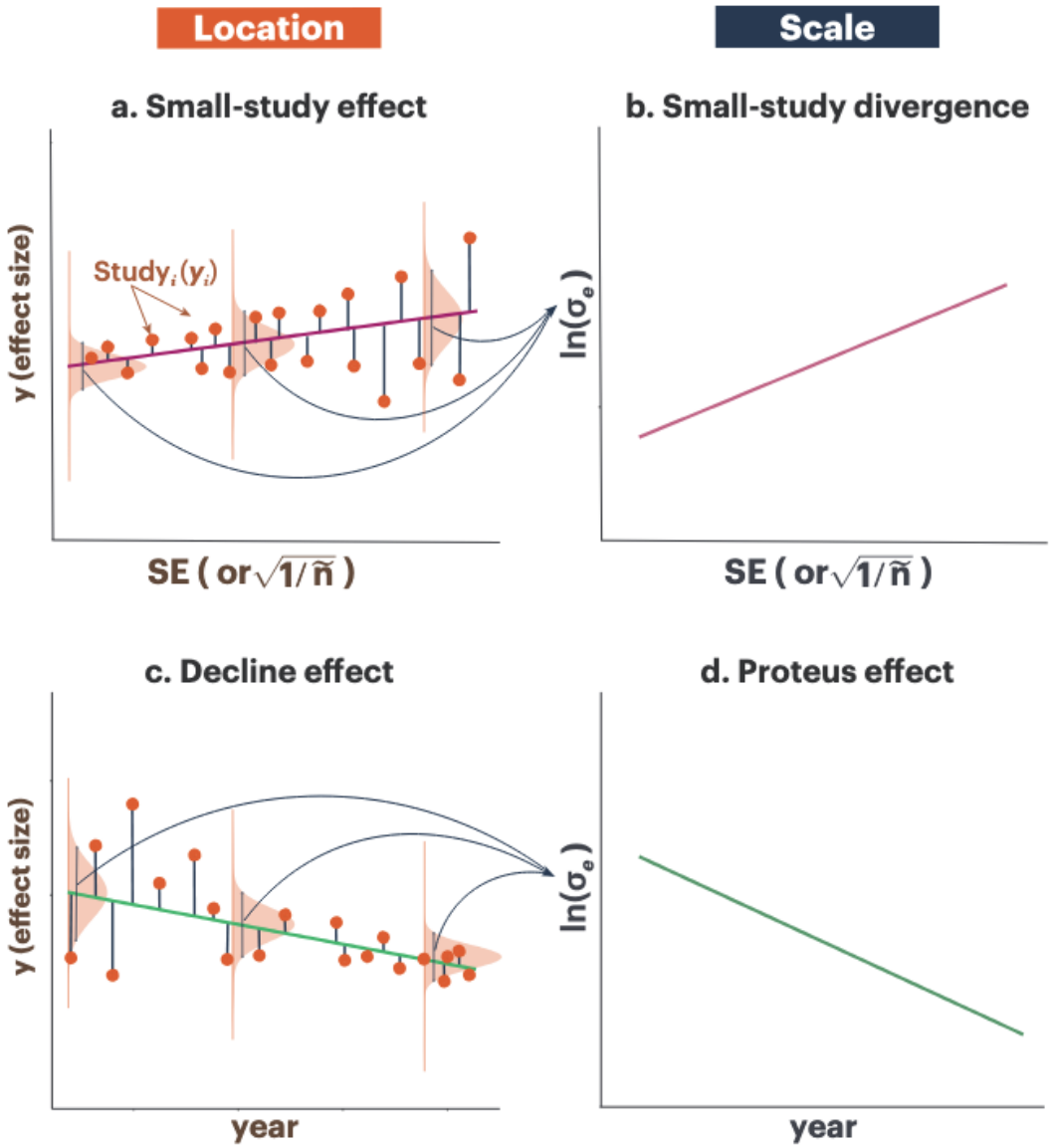

Figure 3: Four types of publication bias: (a) *Small-study effect (location part):* A conventional Egger-type test regresses observed effect sizes (y) on their standard errors (SE or $\sqrt{1/\tilde{n}}$). A significant slope suggests that smaller (less precise) studies yield systematically different mean effects. (b) *Small-study divergence (scale part):* Location-scale models allow testing whether less precise (smaller) studies exhibit not just different average outcomes but also greater (or lesser) variance. (c) *Decline effect (location part):* Also called the time-lag bias, where earlier studies may report inflated effects that gradually decline over publication years (green slope). (d) *Proteus effect (scale part):* Over time, variance among effect sizes could increase or decrease. A decrease might reflect an emerging consensus, whereas an increase may arise if subsequent studies expand across different conditions, species, or methodologies. By including moderators such as sample size or publication year in the scale component, location-scale models can detect biases that inflate variance, revealing more complex patterns of publication distortions beyond mean shifts alone (note all effect sizes are assumed to be independent so one effect size per study).

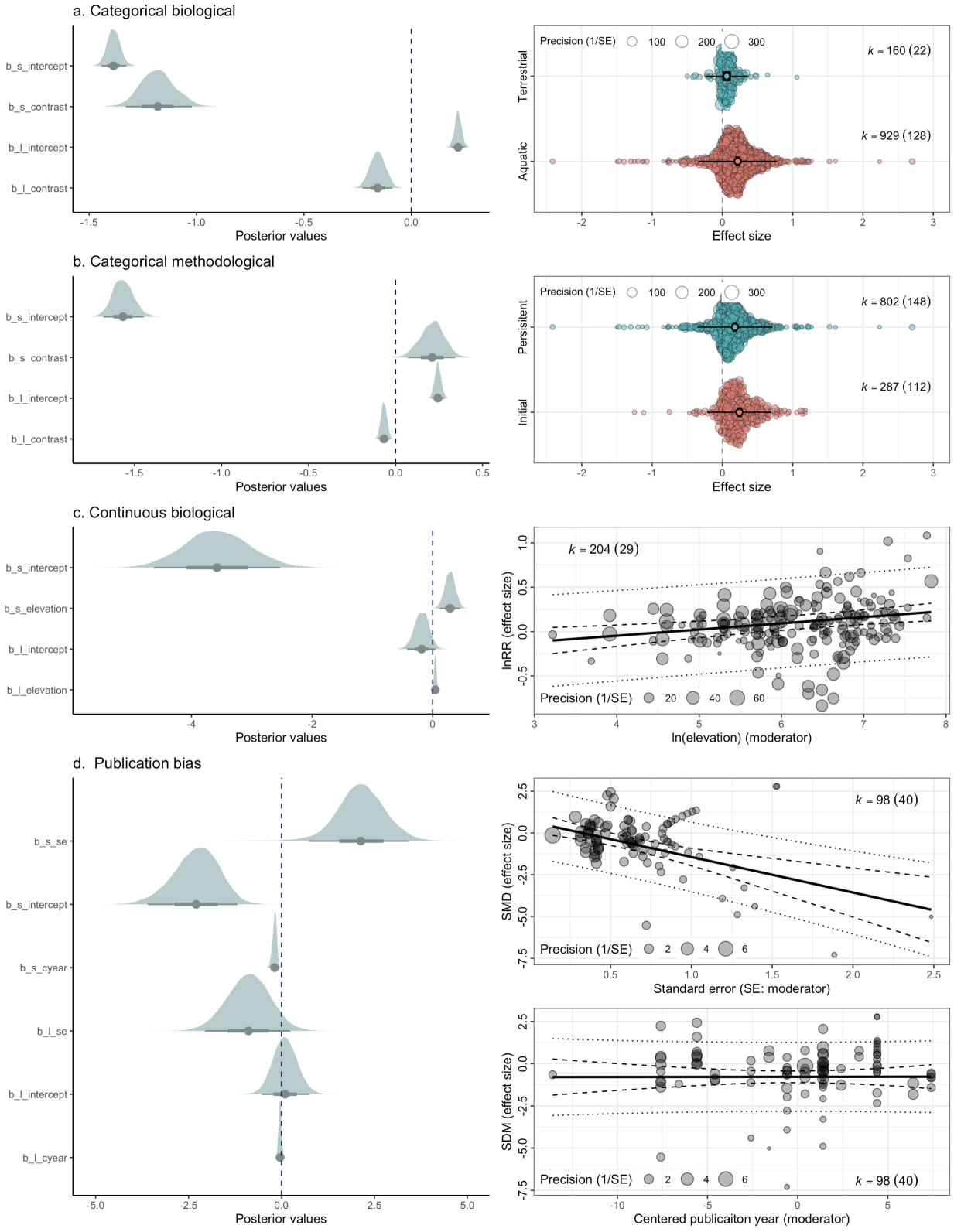

Figure 4: Illustrative location-scale meta-regression examples covering categorical, continuous, and publication bias moderators using 'ggplot2' (Wickham, 2011), 'tidybayes' (Kay, 2020) and 'orchaRd' packages (Nakagawa, Lagisz, O'Dea, et al., 2023) :(see the next page)

(a) *Categorical biological* moderator: contrasting terrestrial (blue) vs. aquatic (red) organisms (terrestrial - aquatic). The left panel shows posterior distributions of four key parameters from the Bayesian location-scale model: the intercept (`b_l_intercept`), the habitat contrast (`b_l_contrast`) in the location part, and the corresponding intercept (`b_s_intercept`) and habitat contrast (`b_s_contrast`) in the scale part. The vertical dashed line indicates zero, aiding the interpretation of effect direction with thick lines showing 66% credible intervals and thin whiskers 95% credible intervals. The right panel or orchard plot depicts effect sizes by habitat (vertical axis) and their average precision (bubble size) or sampling effort (horizontal jitter), illustrating that aquatic organisms showed not only larger mean effect sizes but also higher variance with thick lines showing 95% confidence intervals and thin whiskers 95% prediction intervals. (b) *Categorical methodological* moderator: initial versus persistent temperature exposures (persistent - initial). The left panel similarly displays the posterior distributions for intercept and contrast in both location and scale parts, revealing that persistent exposures yield higher variance than initial exposures. The right panel shows a orchard plot of effect sizes by the method category, with bubble size again proportional to precision (1/SE). (c) *Continuous biological* moderator (e.g., log-elevation). The left panel highlights how both the location (e.g., `b_l_elevation`) and scale (`b_s_elevation`) slopes differ from zero, indicating that mean effect sizes increase with elevation while variance likewise expands. The right panel shows a scatter of effect sizes across the moderator axis (ln(elevation)), with bubble sizes proportional to precision, along with the fitted location trend (solid line) and its 95% confidence intervals (dashed lines) and 95% prediction intervals (dotted lines) (d) *Publication-bias* variables: sample size (SE) and publication year (`cyear`). On the left, the location part (`b_l_se`, `b_l_cyear`) tests for the small-study effect and decline effect (no statistical evidence for these effects), while the scale part (`b_s_se`, `b_s_cyear`) examines small-study divergence and the Proteus effect (evidence for these effects). The right panels illustrate partial regressions against standard error and centred publication year, each bubble sized by precision, with the fitted lines in black and 95% intervals in dashed lines. The bubble plot, based on a standard meta-regression not location-scale meta-regression, for `se` showed a small study effect yet, this effect was not detected in the corresponding location-scale model, this indicates the small-study divergence (which was not modelled) has created the small-study effect in the normal meta-regression, emphasising the importance of all four publication bases as proposed here.

References

Blowes, S. A. (2024). Known unknowns and model selection in ecological evidence synthesis. *bioRxiv*, https://doi.org/10.1101/2024.12.18.629303.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, *80*, 1–28.

Cairns, M., & Prendergast, L. A. (2022). On ratio measures of heterogeneity for meta-analyses. *Research Synthesis Methods*, *13*(1), 28–47.

Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., & Sutherland, W. J. (2019). Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, *56*(12), 2742–2754.

Cinar, O., Nakagawa, S., & Viechtbauer, W. (2022). Phylogenetic multilevel meta-analysis: A simulation study on the importance of modelling the phylogeny. *Methods in Ecology and Evolution*, *13*(2), 383–395.

Cleasby, I. R., & Nakagawa, S. (2011). Neglected biological patterns in the residuals: A behavioural ecologist's guide to co-operating with heteroscedasticity. *Behavioral Ecology and Sociobiology*, *65*, 2361–2372.

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: Statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37.

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach* (Vol. 2). Sage publications.

Dougherty, L. R., & Shuker, D. M. (2015). The effect of experimental design on the measurement of mate choice: A meta-analysis. *Behavioral Ecology*, *26*(2), 311–319.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *bmj*, *315*(7109), 629–634.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature*, *555*(7695), 175–182.

Hadfield, J., & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of evolutionary biology*, *23*(3), 494–508.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological bulletin*, *93*(2), 388.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, *1*(1), 39–65.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, *21*(11), 1539–1558.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557–560.

IntHout, J., Ioannidis, J. P., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: A meta-meta-analysis. *Journal of clinical epidemiology*, *68*(8), 860–869.

Kay, M. (2020). Tidybayes: Tidy data and geoms for bayesian models. *R package version*, *2*(1), 1.

Koricheva, J., & Kulinskaya, E. (2019). Temporal instability of evidence base: A threat to policy making? *Trends in ecology & evolution*, *34*(10), 895–902.

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(4), 619–656.

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society Series C: Applied Statistics*, *55*(2), 139–185.

Lynch, M. (1991). Methods for the analysis of comparative data in evolutionary biology. *Evolution*, *45*(5), 1065–1080.

Mathot, K. J., Arteaga-Torres, J. D., Besson, A., Hawkshaw, D. M., Klappstein, N., McKinnon, R. A., Sridharan, S., & Nakagawa, S. (2024). A systematic review and meta-analysis of unimodal and multimodal predation risk assessment in birds. *Nature Communications*, *15*(1), 4240.

Midolo, G., De Frenne, P., Hölzel, N., & Wellstein, C. (2019). Global patterns of intraspecific leaf trait responses to elevation. *Global Change Biology*, *25*(7), 2485–2498.

Moreno, S. G., Sutton, A. J., Ades, A., Stanley, T. D., Abrams, K. R., Peters, J. L., & Cooper, N. J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. *BMC medical research methodology*, *9*, 1–17.

Mulder, H. A., Gienapp, P., & Visser, M. E. (2016). Genetic variation in variability: Phenotypic variability of fledging weight and its evolution in a songbird population. *Evolution*, *70*(9), 2004–2016.

Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, *13*(1), 4–21.

Nakagawa, S., Lagisz, M., O'Dea, R. E., Pottier, P., Rutkowska, J., Senior, A. M., Yang, Y., & Noble, D. W. (2023). Orchard 2.0: An r package for visualising meta-analyses with orchard plots. *Methods in Ecology and Evolution*, *14*(8), 2003–2010.

Nakagawa, S., Lagisz, M., O'Dea, R. E., Rutkowska, J., Yang, Y., Noble, D. W., & Senior, A. M. (2021). The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. *Research Synthesis Methods*, *12*(1), 4–12.

Nakagawa, S., Noble, D. W., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC biology*, *15*, 1–14.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: Ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6*(2), 143–152.

Nakagawa, S., & Santos, E. S. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, *26*, 1253–1274.

Nakagawa, S., Yang, Y., Macartney, E. L., Spake, R., & Lagisz, M. (2023). Quantitative evidence synthesis: A practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. *Environmental Evidence*, *12*(1), 8.

Neuschulz, E. L., Mueller, T., Schleuning, M., & Böhning-Gaese, K. (2016). Pollination and seed dispersal are the most threatened processes of plant regeneration. *Scientific Reports*, *6*(1), 29839.

Noble, D. W., Lagisz, M., O'dea, R. E., & Nakagawa, S. (2017). Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses.

O'Dea, R. E., Noble, D. W., & Nakagawa, S. (2022). Unifying individual differences in personality, predictability and plasticity: A practical guide. *Methods in Ecology and Evolution*, *13*(2), 278–293.

Pearse, W. D., Davies, T. J., & Wolkovich, E. (2023). How to define, use, and interpret pagel's λ (lambda) in ecology and evolution. *bioRxiv*, 2023–10.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I.-C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., et al. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, *355*(6332), eaai9214.

Pitt, D., Trück, S., van den Honert, R., & Wong, W. W. (2020). Modeling risks from natural hazards with generalized additive models for location, scale and shape. *Journal of Environmental Management*, *275*, 111075.

Pottier, P., Burke, S., Zhang, R. Y., Noble, D. W., Schwanz, L. E., Drobniak, S. M., & Nakagawa, S. (2022). Developmental plasticity in thermal tolerance: Ontogenetic variation, persistence, and future directions. *Ecology Letters*, *25*(10), 2245–2268.

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rodriguez, J. E., Williams, D. R., & Bürkner, P.-C. (2023). Heterogeneous heterogeneity by default: Testing categorical moderators in mixed-effects meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *76*(2), 402–433.

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance-estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, *42*, 1–10.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 1–7.

Sae-Lim, P., Kause, A., Janhunen, M., Vehviläinen, H., Koskinen, H., Gjerde, B., Lillehammer, M., & Mulder, H. A. (2015). Genetic (co) variance of rainbow trout (oncorhynchus mykiss) body weight and its uniformity across production environments. *Genetics Selection Evolution*, *47*, 1–10.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S. A., & Nakagawa, S. (2016). Heterogeneity in ecological and evolutionary meta-analyses: Its magnitude and implications. *Ecology*, *97*(12), 3293–3299. https://doi.org/https://doi.org/10.1002/ecy.1591

Senior, A. M., Viechtbauer, W., & Nakagawa, S. (2020). Revisiting and expanding the meta-analysis of variation: The log coefficient of variation ratio. *Research Synthesis Methods*, *11*(4), 553–567.

Stanley, T., Doucouliagos, H., & Ioannidis, J. P. (2022). Beyond random effects: When small-study findings are more heterogeneous. *Advances in Methods and Practices in Psychological Science*, *5*(4), 25152459221120427.

Takkouche, B., Cadarso-Suarez, C., & Spiegelman, D. (1999). Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *American journal of epidemiology*, *150*(2), 206–215.

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, *189*(4766), 732–735.

Trikalinos, T. A., & Ioannidis, J. P. (2005). Assessing the evolution of effect sizes over time. *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 241–259.

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, *348*(6234), 571–573.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*, 1413–1432.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of statistical software*, *36*(3), 1–48.

Viechtbauer, W., & López-López, J. A. (2022). Location-scale models for meta-analysis. *Research synthesis methods*, *13*(6), 697–715.

Wickham, H. (2011). Ggplot2. *Wiley interdisciplinary reviews: computational statistics*, *3*(2), 180–185.

Williams, D. R., Rodriguez, J. E., & Bürkner, P.-C. (2021). Putting variation into variance: Modeling between-study heterogeneity in meta-analysis. *PsyArXiv*.

Yang, Y., Hillebrand, H., Lagisz, M., Cleasby, I., & Nakagawa, S. (2022). Low statistical power and overestimated anthropogenic impacts, exacerbated by publication bias, dominate field studies in global change biology. *Global Change Biology*, *28*(3), 969–989.

Yang, Y., Lagisz, M., & Nakagawa, S. (2023). Decline effects are rare in ecology: Comment. *Ecology*, *104*(8), e4069.

Yang, Y., Macleod, M., Pan, J., Lagisz, M., & Nakagawa, S. (2023). Advanced methods and implementations for the meta-analyses of animal models: Current practices and future recommendations. *Neuroscience & Biobehavioral Reviews*, *146*, 105016.

Yang, Y., Noble, D. W., Spake, R., Senior, A. M., Lagisz, M., & Nakagawa, S. (2023). A pluralistic framework for measuring and stratifying heterogeneity in meta-analyses. *EcoEvoRxiv*, https://doi.org/10.32942/X2RG7X.

Yang, Y., Sánchez-Tójar, A., O'Dea, R. E., Noble, D. W., Koricheva, J., Jennions, M. D., Parker, T. H., Lagisz, M., & Nakagawa, S. (2023). Publication bias impacts on effect size, statistical power, and magnitude (type m) and sign (type s) errors in ecology and evolutionary biology. *BMC biology*, *21*(1), 71.