One dataset, four meta-analyses: synthesising mean effects, within-population variability, and between-population heterogeneity in ecology

Haley Lacza[1], Ayumi Mizuno[1*], Coralie Williams[2,3], Malgorzata Lagisz[1,2], and Shinichi Nakagawa[1,2*]

[1]Collaboration for Open Science and Synthesis in Ecology and Evolution (COSSEE), 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, New South Wales, Australia
[*]amizuno@ualberta.ca and snakagaw@ualberta.ca

January 29, 2026

Author Statement

Conceptualisation: SN, AM. Formal analysis: HL, AM, SN. Funding Acquisition: SN. Investigation: HL, AM, SN, Methodology: HL, AM, SN, Project administration: HL. Supervision: AM, SN. Validation: AM, SN. Visualisation: AM, SN. Writing - Original Draft: HL, AM, SN. Writing - Review & Editing: All authors.

Conflict of Interest

The authors declare no conflict of interest

Data Availability Statement

All data, scripts and relevant files used for this study can be found at the GitHub repository (https://anonymous.4open.science/dataset-four-meta-analyses-3B43/) and a version of it will be archived at Zenodo (link).

Acknowledgement

SN and AM were supported by the Canada Excellence Research Chair (CERC-2022-00074) and NSERC Discovery Grant (RGPIN-2025-04813). SN thank Elena Gazzea for the initial discussion.

Abstract

1. Ecological syntheses (meta-analysis) usually ask "what is the average effect?", but many ecological questions also depend on whether outcomes become more or less variable and whether effects are predictable across contexts.

2. We show how the same dataset can support a coherent workflow that separates: (i) within-population variability (dispersion among individuals or sampling units inside studies) from (ii) between-population heterogeneity (dispersion among effect sizes across studies), and targets both for mean effects and variability effects.

3. Using the organic versus conventional crop-yield dataset as an illustration, along with an online tutorial, we analyse mean effects with the log response ratio (lnRR; Model 1) and within-population variability with the log variance ratio (lnVR) and the log coefficient of variation ratio (lnCVR; Model 2), noting that these three effect sizes can be computed from the same summary statistics (means, SDs and sample sizes).

4. We then extend standard meta-regression to location-scale (mean-variance) modelling, allowing moderators to explain not only how lnRR (Model 3) and lnVR/lnCVR (Model 4) shift on average ("location") but also how their within-study/residual heterogeneity changes with context ("scale"), thereby distinguishing settings where effects are generalisable and transferable from those where they are strongly context-dependent.

5. The core message is that many ecological datasets already contain sufficient information to synthesise performance (lnRR), reliability/stability (lnVR/lnCVR), and predictability (context-dependent heterogeneity; i.e., four models or meta-analyses) side by side. Doing so improves not only statistical inference but also our understanding of the changing world, making meta-analytic outputs and insights more directly decision-relevant.

Keywords: lnRR; lnVR; lnCVR; yield stability; heterogeneity; location–scale models; meta-regression; orchard plots

1 Introduction

Ecological effects vary. The same ecological process or intervention, for example, can produce different outcomes across individuals, populations, species, and environments, and this variation is often large enough to alter statistical inference and to limit generalisation across contexts (Levin, 1992; Lawton, 1999; Chamberlain et al., 2014). This means that ecology may not be well served by questions that focus only on "the average effect". Two further questions are routinely important: "how variable are outcomes?" and "how consistent are effects across contexts?" (Duncan and Kefford, 2021; Spake et al., 2023). These questions are related, but they operate at two different levels under a meta-analytic context. Outcomes can vary more or less within a population (inter-individual variability; e.g., a treatment may benefit some individuals but harm others, widening the spread even if the mean changes little), and effect sizes can vary more or less across studies or populations (between-study heterogeneity; e.g., the same intervention shows a strong effect in some studies but a weak effect elsewhere). Both levels matter for ecological inference and practical application, where generalisation and predictability, whether a phenomenon or effect is observed consistently across different conditions or only under specific conditions, often determine whether evidence is useful. Therefore, meta-analysis should target both the mean effect (central tendency or "location") and variation effect (dispersion or "scale"), rather than treating variation as a nuisance to be averaged away (Nakagawa et al., 2015; Senior et al., 2020; Viechtbauer and López-López, 2022; Nakagawa et al., 2025a).

Most ecological meta-analyses, however, are still mean-centric. A typical synthesis estimates an overall mean effect (e.g., using the log-response ratio, lnRR, or the standardised mean difference, d) (Hedges et al., 1999; Lajeunesse, 2011, 2015). Variation enters mainly in two limited ways. First, within-study variation (standard deviations) is usually treated as input for sampling variances and weights, rather than as a biological outcome that might itself change with treatments or environments (e.g., an intervention might stabilise outcomes by reducing among-individual differences, or destabilise them by amplifying sensitivity). Second, variation among effect sizes is summarised as "heterogeneity", reported as I^2 or as a random-effects variance component (often written as τ^2) (Higgins and Thompson, 2002; Higgins et al., 2003), and commonly assumed to be constant across studies (i.e., "homoscedasticity"; in contrast, non-constant variance across a continuous variable or among groups are called "heteroscedasticity"). Measuring heterogeneity in a meta-analysis is informative, but simply reporting it does not show how predictable or generalisable the average effect is across contexts. What is often more important is understanding when and why effect sizes vary between studies. For example, if effect sizes are similar in some settings but highly variable in others, such as under different ecological conditions or study designs, then relying on an overall average effect may be a poor way to predict outcomes in a new context.

Two complementary method developments allow us to extract much more from the same dataset (i.e., evidence base). First, "meta-analysis of variation" uses dispersion-based effect sizes such as the log variance ratio (lnVR) and the log coefficient of variation ratio (lnCVR) to quantify changes in within-population dispersion (inter-individual variability)

between comparison groups (e.g., heteroscedasticity between the control and treatment groups) (Nakagawa et al., 2015; Senior et al., 2020) (Fig. 1A–C). A major advantage is practical: these effect sizes can usually be computed from the same summary statistics already collected for mean-based synthesis (means, standard deviations, and sample sizes). Second, "location-scale" meta-regression extends standard meta-regression models by allowing the amount of among-effect-size variability to depend on moderators, thereby testing whether heterogeneity can differ between different moderator groups or contexts (i.e., heteroscedasticity among moderator levels) (Viechtbauer and López-López, 2022; Nakagawa et al., 2025a). Throughout this paper, we use "variability" to mean within-study/within-population dispersion (targeted by lnVR/lnCVR), and "heterogeneity" to mean dispersion among effect sizes across studies, which can be quantified for both mean effects (e.g., lnRR or d) and variability effects (lnVR/lnCVR) via variance components or moderators in meta-analyses and meta-regression analyses (Fig. 1E–H).

Here, we illustrate "one dataset, four meta-analyses" using the organic versus conventional crop-yield dataset compiled by Ponisio et al. (2015). We begin with the standard mean-effect synthesis (lnRR) to summarise average yield differences (Model 1). We then analyse lnVR/lnCVR to address a different but equally ecological question: namely whether organic systems differ from conventional systems in within population yield variability (i.e., stability) (Model 2), a question highlighted in Knapp and van der Heijden (2018). Next, we move from description to explanation by fitting two types of meta-regression models: (i) location (mean) meta-regression models to test which moderators shift average effects, and (ii) location-scale (mean-variance) models to test whether the amount of between-study heterogeneity changes systematically with those moderators (Models 3 & 4) (Viechtbauer and López-López, 2022; Nakagawa et al., 2025a). Note that Model 3 applies the location-scale framework to mean effects (lnRR), whereas Model 4 applies the same framework to variability effects (lnVR/lnCVR). We conclude with practical guidance on effect-size choice and reporting. We explain how to clearly distinguish changes in within-population stability from variation in effect sizes among studies, and why it is important to retain and report the basic summary statistics (means, standard deviations, and sample sizes) needed to synthesise both mean effects and variability effects. Importantly, our illustrative examples are implemented in R using `metafor` (Viechtbauer, 2010), `brms` (Bürkner, 2017) and `glmmTMB` (Brooks et al., 2017; Kristensen et al., 2026), accessible via an online tutorial.

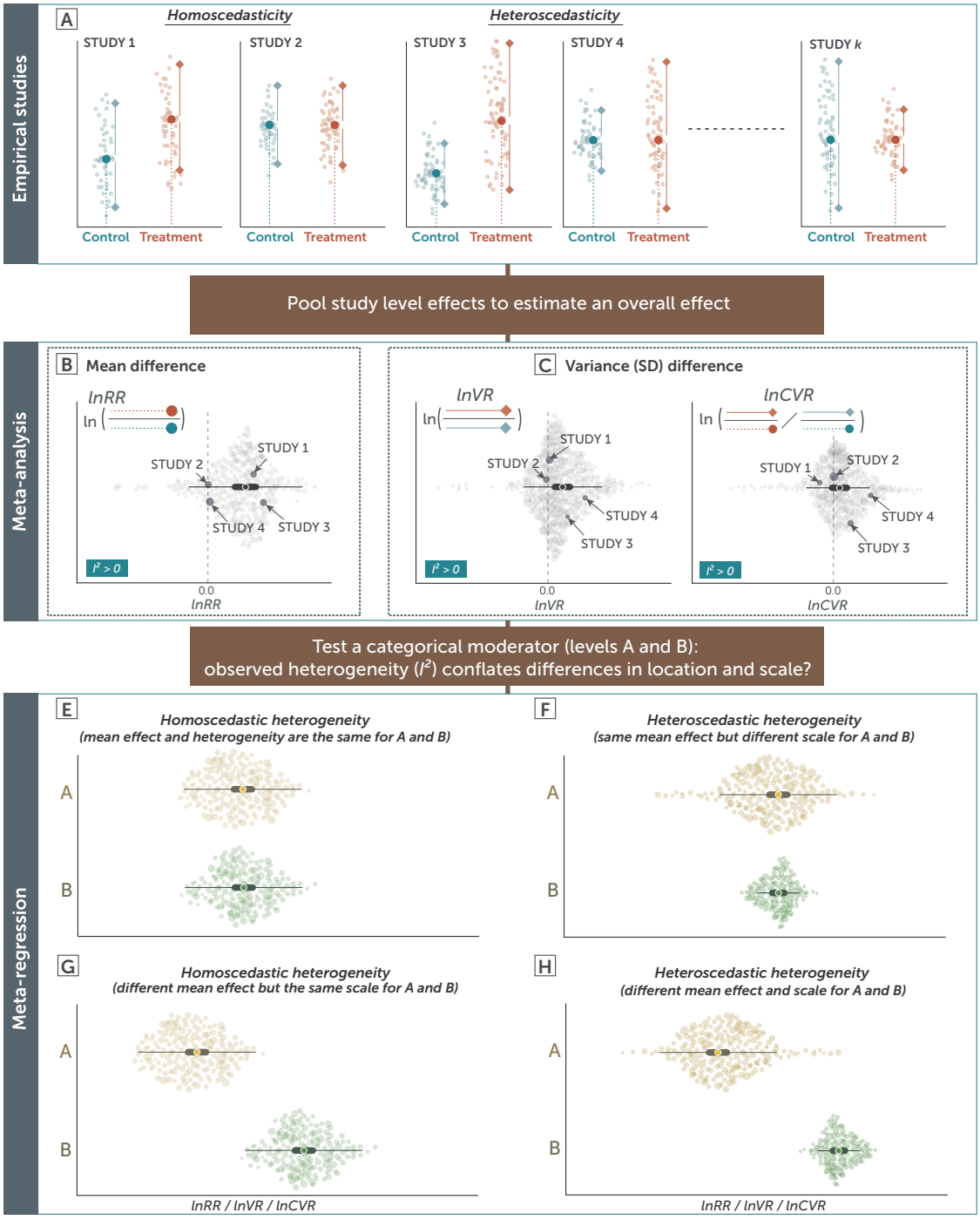

Figure 1: **Conceptual overview of synthesising mean effects, variability effects, and location-scale modelling of heterogeneity from empirical studies**

Figure 1 (continued). (A) Empirical two-group studies (control vs treatment) can be homoscedastic (similar within-group dispersion; Studies 1-2) or heteroscedastic (different within-group dispersion; Studies 3–4). Each study yields a study-level effect size that targets either a change in location (mean) and/or a change in scale (dispersion). (B) Mean effects are summarised with a mean-based effect size such as the log response ratio, lnRR = $\ln(\bar{X}_T/\bar{X}_C)$, and then pooled across studies to estimate an overall mean effect (Model 2). (C) Within-population variability (dispersion) effects are summarised with dispersion-based effect sizes such as the log variance ratio (lnVR = $\ln(s_T/s_C)$) and/or the log coefficient of variation ratio (lnCVR = $\ln[(s_T/\bar{X}_T)/(s_C/\bar{X}_C)]$), and then pooled across studies to estimate an overall variability effect (Model 2). In (B–C), each point represents a study-level effect size; even when the pooled mean is near zero (vertical dashed line), effect sizes can vary substantially across studies, producing between-study heterogeneity ($I^2 > 0$). (E–H) Location–scale (mean–variance) meta-regression tests whether a categorical moderator (levels A and B) shifts the expected effect (location) and/or the predictability/transferability of effects (scale; heterogeneity) (Models 3 and 4). (E) Homoscedastic heterogeneity: mean effect and heterogeneity are the same for A and B. (F) Heteroscedastic heterogeneity: mean effects are the same, but heterogeneity differs between A and B. (G) Homoscedastic heterogeneity: mean effects differ between A and B, but heterogeneity is the same. (H) Heteroscedastic heterogeneity: both mean effects and heterogeneity differ between A and B.

2 Meta-analysis of mean

2.1 Why lnRR remains a useful starting point

Mean-effect meta-analysis is usually where ecological synthesis begins: "what is the average effect?" For two-group comparisons, a common choice is the log response ratio (lnRR), because it represents proportional change, is interpretable on a multiplicative scale, and is widely used across ecological meta-analyses (Hedges et al., 1999; Lajeunesse, 2011, 2015). For each comparison (treatment T vs control C), where i indexes the effect size for an individual treatment–control comparison, let $\bar{X}_{g,i}$ denote the sample mean, $s_{g,i}$ the sample standard deviation, and $n_{g,i}$ the sample size in group $g \in \{T, C\}$. The lnRR point estimate is:

$$y_i^{(\mathrm{lnRR})} = \ln\left(\frac{\bar{X}_{T,i}}{\bar{X}_{C,i}}\right), \tag{1}$$

and its sampling variance is written as $v_i^{(\mathrm{lnRR})}$. Under the standard assumption of independent groups, a common approximation is

$$v_i^{(\mathrm{lnRR})} \approx \frac{s_{T,i}^2}{n_{T,i}\,\bar{X}_{T,i}^2} + \frac{s_{C,i}^2}{n_{C,i}\,\bar{X}_{C,i}^2}, \tag{2}$$

so each effect size enters the meta-analysis with its own known (approximated) plug-in precision (Hedges et al., 1999; Lajeunesse, 2010, 2015). We use v_i throughout as shorthand for "the sampling variance of effect size i", with the

superscript indicating which effect size is being analysed (see also Senior et al., 2020; Nakagawa et al., 2023b).

lnRR is a practical starting point because the ingredients needed to compute it (means, SDs, and sample sizes) are typically the same summary statistics required to compute lnVR and lnCVR. In other words, many existing datasets already contain what is needed to analyse mean effects and within-population variability side-by-side (Nakagawa et al., 2015; Senior et al., 2020). We do not focus on the standardised mean difference (d) in this article because ratio-scale outcomes are common in ecology (making lnRR a natural scale), and because the broader message of this paper is about treating differences in variation as targets of synthesis rather than as background assumptions (Nakagawa et al., 2015; Senior et al., 2020).

2.2 Multilevel meta-analysis as the default in ecology

A recurring issue in ecological meta-analysis is treating effect sizes as independent, recently highlighted by Peacor et al. (2025). In practice, multiple effect sizes commonly come from the same paper (e.g., different species, sites, years, outcomes, or treatment contrasts), and additional dependence can arise through shared controls or repeated measurements (Lajeunesse, 2011; Nakagawa and Santos, 2012). For this reason, a multilevel (random-effects) model is a sensible default, as it separates variation across different levels and avoids overstating precision (i.e., reducing inflated Type I error rates).

Using the notation adopted in the past work Nakagawa and Santos (2012), a simple three-level model can be written as:

$$y_i = \beta_0 + u_{j[i]} + e_i + m_i, \tag{3}$$

$$u_j \sim \mathcal{N}(0, \sigma_u^2), \tag{4}$$

$$e_i \sim \mathcal{N}(0, \sigma_e^2), \tag{5}$$

$$m_i \sim \mathcal{N}(0, v_i), \tag{6}$$

where y_i is the observed effect size (here, $y_i^{(\mathrm{lnRR})}$), β_0 is the intercept (overall or meta-analytic mean), $u_{j[i]}$ captures between-study (or between-paper) differences, e_i captures within-study differences among multiple effect sizes from the same study, and m_i represents sampling error with known sampling variance v_i (Lajeunesse, 2011; Nakagawa et al., 2025a). In practice, this model matches how ecological evidence is generated: individual studies (papers) often contribute clusters of related comparisons, and this dependence should be modelled rather than ignored. When sampling errors are correlated, for example, because multiple effect sizes share a control group, m_i can be generalised to allow covariances. However, for notational simplicity, we retain the scalar form v_i (assuming independence among sampling errors) rather than the full variance–covariance matrix representation (Williams et al., 2025).

This model is a multilevel random-effects meta-analysis, in which true effect sizes are allowed to vary across studies and comparisons rather than being assumed to share a single common value. This enables variance decomposition and makes explicit what is meant by "heterogeneity" (variance not due to sampling error) and how it can be summarised. In simple random-effects meta-analyses (with a single study-level random effect in addition to sampling error), the total heterogeneity is often denoted by τ^2. In the multilevel model considered here, this total heterogeneity is partitioned into a between-study component (σ_u^2) and a within-study component (σ_e^2), whose sum corresponds to τ^2. A common descriptive summary of heterogeneity is I^2, the proportion of the total variance in observed effect sizes that is attributable to heterogeneity rather than sampling error (Lajeunesse, 2011). For the multilevel model in Equation (3), a convenient definition of the total heterogeneity, equivalent to standard I^2 in the meta-analysis literature (Higgins and Thompson, 2002; Higgins et al., 2003), is:

$$I^2_{\text{total}} = \frac{\sigma_u^2 + \sigma_e^2}{\sigma_u^2 + \sigma_e^2 + \bar{v}}, \tag{7}$$

where $\bar{v}$ is a representative sampling variance (the mean or similar of v_i across effect sizes) (Nakagawa and Santos, 2012). The same decomposition yields component-specific contributions,

$$I^2_{\text{between}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2 + \bar{v}}, \tag{8}$$

$$I^2_{\text{within}} = \frac{\sigma_e^2}{\sigma_u^2 + \sigma_e^2 + \bar{v}}, \tag{9}$$

which indicate whether most heterogeneity sits between studies or among effect sizes within studies. These summaries are useful diagnostics, but they remain descriptive: Because I^2 is context-dependent, large values (total or component-specific) indicate that effects vary across contexts and motivate the next step, testing whether moderators explain heterogeneity in a meta-regression (Yang et al., 2025) which we demonstrate in Section 4.

2.3 Example: organic versus conventional farming average yield

We used a part of the organic versus conventional crop-yield database compiled by Ponisio et al. (2015); the dataset included 318 comparisons (effect sizes) for cereal crops (e.g., maize, wheat, barley, and oats), where organic systems were compared against conventional farming systems using standard synthetic inputs (control). Importantly, the type of fertiliser (animal- vs plant-derived) varied only within organic systems. We used this fertiliser-type variable as a moderator of the organic–conventional contrast, testing whether the magnitude of the organic–conventional effect varied with organic management. We fitted the three level meta-analysis described above with study ID as a random effect (between-study effects) and effect (size) ID as another random effect (within-study effects); note that our model is for an illustrative purposes, and, therefore, we do not claim this is the best meta-analytic model (e.g., an alternative model could include another random effect for cereal types; for more detail, see the online tutorial). Our

mean-effect synthesis based on lnRR answers a clear question (Fig. 1B): on average, how do organic and conventional yields compare? As with Ponisio et al. (2015), we found a clear decline in the average yield for organic farming by approximately 27% (Model 1 using `metafor`: β_0 = -0.32,$1 - \exp(-0.32) \approx 0.27$, 95% confidence interval, CI = [-0.43, -0.22]; Fig. 2A,C). Yet, there was high heterogeneity ($I^2_{\text{total}} = 98.11$, $I^2_{\text{between}} = 46.09$, and, $I^2_{\text{within}} = 52.02$; Fig. 2E), which warrants meta-regression analysis (see Section 4; for comparable models using `glmmTMB` and `brms`, see the online tutorial).

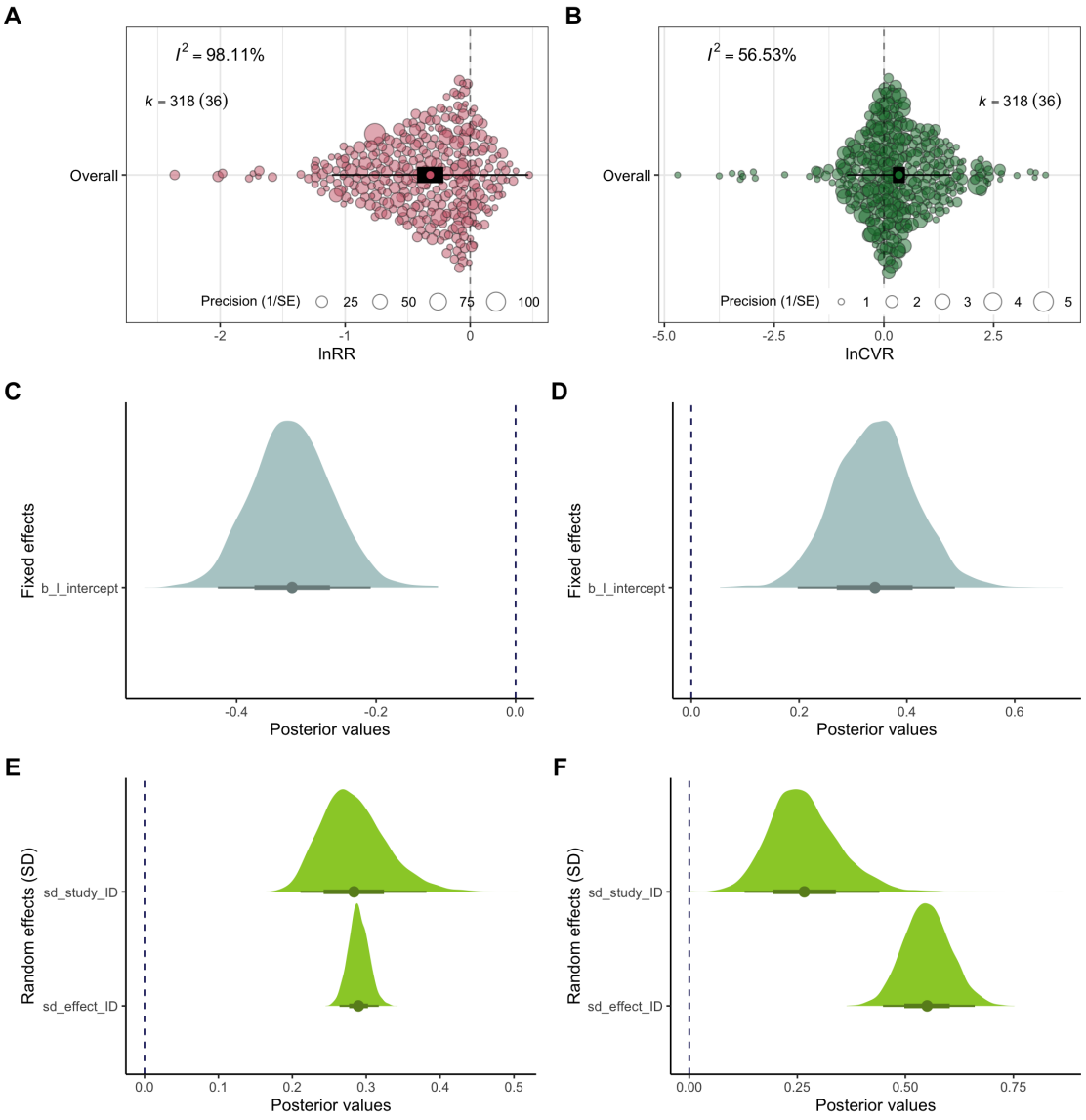

Figure 2: **Meta-analyses of mean yield differences and yield stability from a single evidence base.** Panels (A,C,E) summarise the meta-analysis of mean effects using the log response ratio (lnRR), and panels (B,D,F) summarise the meta-analysis of within-population (mean-corrected) variability using the log coefficient of variation ratio (lnCVR), both comparing organic to conventional cereal yields (Ponisio dataset; k effect sizes with the number of studies in parentheses). (A,B) Orchard plots show individual effect sizes (points; size proportional to precision, 1/SE) with horizontal jitter to reduce overplotting. The black square denotes the pooled estimate from a multilevel random-effects model fitted using `metafor`; thick horizontal bars indicate the 95% interval for the pooled mean, and thin whiskers indicate the 95% prediction interval. The vertical dashed line marks zero (no organic–conventional difference). The reported I^2 gives the proportion of total variation attributable to heterogeneity rather than sampling error. (C,D) Posterior distributions for the overall (intercept) effect in each model (b_0 = `b_l_intercept`), estimated using Bayesian multilevel models fitted in `brms`, where negative lnRR implies lower organic yield and positive lnCVR implies higher relative variability (lower stability) under organic management. (E,F) Posterior distributions for random-effects standard deviations (study ID and effect-size ID). Points are posterior medians; thick and thin intervals denote 66% and 95% credible intervals, respectively. The `R` packages: `ggplot2` (Wickham, 2011), `tidybayes` (Kay, 2020) and `orchaRd` (Nakagawa et al., 2023a) were used to draw these plots.

3 Meta-analysis of variability

3.1 Why variability deserves its own meta-analysis

Mean effects describe shifts in central tendency, but many ecological questions are also about "spread": do outcomes become more or less variable within populations when conditions change? Within-population variability is not simply nuisance variation. It can reflect differential sensitivity among individuals, temporal or spatial environmental variability, changes in trait distributions, or shifting demographic and community composition. Importantly, variability can change even when mean outcomes do not (Sánchez-Tójar et al., 2020; Senior et al., 2016a).

Despite this, standard meta-analytic workflows usually treat within-study standard deviations mainly as inputs for sampling variances and weights (as in Equation (2)). As a result, the same dataset can be used to make strong statements about mean effects while leaving within-population stability unanalysed. Meta-analysis of variability addresses this gap by analysing effect sizes that directly compare dispersion between groups, allowing the same evidence base to synthesise both mean outcomes and within-population variability (Nakagawa et al., 2015; Senior et al., 2020).

3.2 Effect sizes for within-population variability: lnVR and lnCVR

For two-group designs (treatment T vs control C), within-population variability can be compared using two closely related dispersion-based effect sizes (Nakagawa et al., 2015; Senior et al., 2020). The log variance ratio (lnVR) targets differences in absolute dispersion (on the SD scale):

$$y_i^{(\mathrm{lnVR})} = \ln\left(\frac{s_{T,i}}{s_{C,i}}\right), \tag{10}$$

so $y_i^{(\mathrm{lnVR})} > 0$ indicates higher within-population variability under treatment. lnVR is most natural when the scientific question is about absolute spread on the measurement scale (e.g., whether an intervention increases or decreases variability regardless of any shift in the mean). For independent groups, a commonly used large-sample approximation to the sampling variance is

$$v_i^{(\mathrm{lnVR})} \approx \frac{1}{2\left(n_{T,i}-1\right)} + \frac{1}{2\left(n_{C,i}-1\right)}, \tag{11}$$

with small-sample bias corrections available and recommended when sample sizes are modest (Nakagawa et al., 2015; Senior et al., 2020).

The log coefficient of variation ratio (lnCVR) targets dispersion relative to the mean:

$$y_i^{(\mathrm{lnCVR})} = \ln\left(\frac{s_{T,i}/\bar{X}_{T,i}}{s_{C,i}/\bar{X}_{C,i}}\right), \tag{12}$$

and is often preferred in ecology because standard deviations frequently scale with means (including for yields, abundances, and performance measures) (Nakagawa et al., 2015; Senior et al., 2020). Put simply, lnVR asks whether the SD changes, whereas lnCVR asks whether the SD changes more (or less) than the mean. For independent groups, a convenient approximation to the sampling variance combines the lnVR component with the additional uncertainty from the means:

$$v_i^{(\mathrm{lnCVR})} \approx \frac{1}{2\,(n_{T,i}-1)} + \frac{1}{2\,(n_{C,i}-1)} + \frac{s_{T,i}^2}{n_{T,i}\,\bar{X}_{T,i}^2} + \frac{s_{C,i}^2}{n_{C,i}\,\bar{X}_{C,i}^2}, \tag{13}$$

again with small-sample bias corrections and extensions for dependent designs (e.g., paired or pre-post data) given by Senior et al. (2020).

A major practical advantage is that both lnVR and lnCVR can usually be computed from the same summary statistics already extracted for lnRR (means, SDs, and sample sizes) (Nakagawa et al., 2015; Senior et al., 2020). Once calculated, they can be analysed using the same multilevel random-effects framework introduced for mean effects (Equation 3), simply replacing $y_i^{(\mathrm{lnRR})}$ with $y_i^{(\mathrm{lnVR})}$ or $y_i^{(\mathrm{lnCVR})}$ and using the corresponding known v_i. The question changes focus from mean to focus on within-population variability, but the meta-analytic model stays the same (see Box 1). It would be worth mentioning that you cannot calculate these dispersion-based effect sizes if the within-group variability is not reported, such as when SDs are missing or zero. For lnCVR, you also cannot calculate it if group means are zero or nearly zero.

Box 1 — Meta-analyses of variability across fields: what lnVR and lnCVR add

Meta-analyses usually synthesise differences in group means, but many applied and mechanistic questions hinge on whether an intervention (or condition) changes dispersion within groups, whether outcomes become more homogeneous or more heterogeneous. For example, the logic has been used to address questions that are naturally about heterogeneity among individuals. In nutrition/health, for example, variance-focused meta-analysis has been used to compare dietary interventions not only in terms of average weight-related outcomes, but also in whether they compress or spread the distribution of responses (i.e., whether benefits are broadly shared or concentrated in a subset) (Senior et al., 2016a).

In neuroscience/psychiatry, Brugger et al. (2020) used lnVR and lnCVR to test the hypothesis that schizophrenia is associated with increased inter-individual variability in striatal dopamine function, meta-analysing dozens of studies and finding greater variability in dopamine receptor/transporter availability in patients than in controls. More broadly, the same "variance as an outcome" framing has been applied to clinical efficacy questions, such as whether antipsychotic treatment effects differ in heterogeneity across patient groups, where the decision-relevant issue is not only the expected response but also how widely responses vary (Howes and Chapman, 2024).

In the social sciences, O'Dea et al. (2018) applied lnCVR at a very large scale (over 1.6 million students) to quantify differences in variability of academic grades between boys and girls, showing how dispersion meta-analysis can be used to test claims that are explicitly about variability rather than mean performance. Across these examples, the general idea is the same: synthesising lnVR/lnCVR alongside mean effects shifts the inferential target from "what happens on average?" to "how different do individuals respond?, and how generalisable is it?"

3.3 Example: yield variability (stability) under organic versus conventional farming

A meta-analysis of variability addresses a different ecological question: do organic systems differ in yield variability (i.e., stability) relative to conventional systems (Fig. 1C)? This question has been explicitly raised in the organic-yield literature, where lnCVR, along with lnVR, has been used to quantify relative yield variability and to connect synthesis to concerns about reliability and stability (Knapp and van der Heijden, 2018). Using lnCVR is particularly natural here because yield variability often scales with mean yield; lnCVR therefore targets stability relative to average performance (Knapp and van der Heijden, 2018). As with the previous meta-analysis, there was an increase in (within-population) relative variability in crop yield for organic farming by approximately 41% (Model 2 using `metafor`: $\beta_0 = 0.34$, $\exp(0.34) - 1 \approx 0.41$, 95% confidence interval, CI = [0.20, 0.48]; Fig. 2B,D). Also we found

moderate heterogeneity (`metafor`: $I^2_{\text{total}} = 56.53$, $I^2_{\text{between}} = 9.52$, and, $I^2_{\text{within}} = 47.01$; Fig. 2F); this result indicated we require a further meta-regression analysis (see Section 4; for comparable models using `glmmTMB` and `brms`, see the online tutorial).

4 Location-scale meta-regression models

4.1 From explaining average effects to explaining predictability

Meta-regression is usually motivated by the desire to explain why mean effects differ among studies: which ecological or methodological features shift the expected lnRR (yield gap) or lnCVR (reliability/stability gap)? However, in (applied) ecology, the more consequential question is often whether an effect is predictable in a new setting. Two management options can have similar mean effects yet differ sharply in how consistently those effects appear across studies. Standard multilevel random-effects models acknowledge heterogeneity, but they typically assume that heterogeneity is constant across studies and moderator levels (homoscedasticity). Location-scale models relax this assumption by letting the "amount of heterogeneity" itself depend on context, turning "it depends" from a post-hoc caveat into an estimable, testable component of the synthesis (Duncan and Kefford, 2021; Spake et al., 2023; Viechtbauer and López-López, 2022; Nakagawa et al., 2025a). This is the key step linking synthesis to generalisability and transferability: it tells us not only what to expect on average, but also in which contexts that expectation is reliable enough to apply to a new site or study (Spake et al., 2022, 2023).

4.2 Location-scale meta-regression: letting heterogeneity depend on context

A useful feature of location-scale modelling is that the same framework can be applied to different effect-size streams. In what follows, y_i can represent a mean effect size (e.g., $y_i^{(\text{lnRR})}$) or a variability effect size (e.g., $y_i^{(\text{lnVR})}$ or $y_i^{(\text{lnCVR})}$). The sampling term m_i is paired with the corresponding sampling variance v_i for that effect size (e.g., $v_i^{(\text{lnRR})}$, $v_i^{(\text{lnVR})}$, or $v_i^{(\text{lnCVR})}$). "Location" therefore refers to the expected value of the chosen effect size, while "scale" refers to the residual (within-study) heterogeneity in that effect size and how it changes with context. This allows moderators to be evaluated for two distinct roles: shifting the expected effect and altering its predictability (uncertainty around the average) across studies.

A multilevel location (mean) meta-regression can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + u_{j[i]} + e_i + m_i, \tag{14}$$

where $x_{1i}, \ldots, x_{pi}$ are the values of p moderator variables for effect size i and β_p are their corresponding regression coefficients; the β's describe how moderators shift the average effect (whether the outcome is lnRR, lnVR, or lnCVR). This meta-regression model improves interpretation because it turns an overall average into a context-conditional average when a meta-analytic model (Equation (3)) finds non-zero heterogeneity, which is almost always the case for ecological datasets (Senior et al., 2016b).

Location–scale meta-regression extends Equation (14) by allowing the residual (within-study) heterogeneity to vary with moderators. One convenient formulation is a double-hierarchical model in which the effect-size-level variance is modelled on the log scale:

$$y_i = \beta_0^{(l)} + \beta_1^{(l)} x_{1i} + \cdots + \beta_p^{(l)} x_{pi} + u_{j[i]}^{(l)} + e_i^{(l)} + m_i, \quad \text{(location sub-model)} \tag{15}$$

$$e_i^{(l)} \sim \mathcal{N}\left(0, \sigma_{e,i}^2\right), \tag{16}$$

$$\ln(\sigma_{e,i}) = \beta_0^{(s)} + \beta_1^{(s)} x_{1i} + \cdots + \beta_p^{(s)} x_{pi}. \quad \text{(scale sub-model)} \tag{17}$$

Here, the superscripts (l) and (s) denote the location (mean) and scale (variance, or more precisely, the natural logarithm of standard deviation) components; note that `metafor` models $\ln(\sigma^2)$ whereas `brms` and `glmmTMB` model $\ln(\sigma)$. A non-zero scale coefficient ($\beta_p^{(s)} \neq 0$) implies that the amount of heterogeneity depends on the moderator: effects are more variable in some contexts than others, so generalisation is correspondingly easier or harder (Viechtbauer and López-López, 2022; Nakagawa et al., 2025a). It is noted that Nakagawa et al. (2025a) introduces more complex location-scale models with the between-study effect on the scale part (i.e., $u_{j[i]}^{(l)}$), and an even more complex one with a correlation (ρ) modelled between $u_{j[i]}^{(s)}$ and $u_{j[i]}^{(l)}$; a positive correlation (ρ) indicates, for example, larger effect sizes tend to have larger deviations (for these models, see the online tutorial).

Importantly, a moderator in a location-scale model can matter in two qualitatively different ways. It can shift the average effect (captured in the location part), but it can also change the spread of effects across studies (captured in the scale part). The second case is often what readers mean by "context dependence": the expected effect may not change much, but the evidence becomes more or less predictable (see Box 2). This very point is well illustrated in the following example.

Box 2 — Location–scale modelling as a general tool for discovering heteroscedasticity, not just meta-analysis

A recurring limitation of "mean-only" modelling is that it treats heteroscedasticity as an inconvenience: if residual variance differs among individuals, sites, years, taxa, or experimental settings, it is typically absorbed into a single error term and reported (at best) as overdispersion or "unexplained variability" (Nakagawa et al., 2025a,b,c). Location-scale modelling flips this perspective by treating variability as an outcome in its own right. The conceptual move is simple: fit a model for the mean (location) in the usual way, and then fit a second model for the dispersion (scale), often by modelling $\log(\sigma)$ or $\log(\sigma^2)$ as a linear predictor with its own covariates. This allows the same predictors to be assessed for two distinct roles: whether they shift the expected response, and whether they change predictability (the spread of responses around that expectation).

Because the scale component is interpretable, location-scale models directly address questions that are intrinsically about individual differences and context dependence. In behavioural ecology, for example, the "personality-predictability" framing asks whether individuals differ not only in average behaviour but also in within-individual variability, and whether those two components covary (Westneat et al., 2013; Cleasby et al., 2015; O'Dea et al., 2022). In quantitative genetics and comparative work, the same logic allows variance components (and their covariate dependence) to be treated as biologically meaningful traits, rather than as a single nuisance parameter (Mulder et al., 2008; Hill and Mulder, 2010; Mulder et al., 2016; Sae-Lim et al., 2015; Nakagawa et al., 2025b). In primary research, this is often exactly what we care about: whether environmental stressors amplify inter-individual differences; whether management interventions stabilise outcomes; whether phenotypes become more canalised in some contexts but more labile in others; or whether treatment effects are reliable enough to generalise.

Crucially, the location-scale idea is not limited to Gaussian-distributed data. Modern "distributional-regression" implementations extend the same two-part thinking to many common outcome types by allowing separate sub-models for multiple distributional parameters, such as "shape" sub-models (e.g., skewness and kurtosis) and zero- and one-inflated sub-models (Lee and Nelder, 1996, 2006; Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2008). For proportions, dispersion (or overdispersion) can be modelled alongside the mean on binomial/beta-type scales; for counts, predictors can act on the mean rate while a separate component captures extra-Poisson variation (e.g., negative-binomial dispersion) and, where relevant, zero inflation (Nakagawa et al., 2025c). This matters in ecology because the empirical signature of "it depends" is frequently variance, not mean: an ecological driver may have only a modest effect on the expected proportion or expected count, yet strongly alter dispersion across sites, seasons, or taxa, signalling that the process is conditional on unmeasured context (cf. Duncan and Kefford, 2021; Spake et al., 2023).

4.3 Example: organic fertiliser type shapes both the yield gap and stability

In the organic-conventional crop-yield dataset (Ponisio et al., 2015), organic fertiliser type (animal- vs plant-derived) is agronomically meaningful and plausibly linked to context-dependence; this single moderator is used in both the location and scale components of the model to test shifts in the expected effect size and in residual heterogeneity (see Fig. 1E–H).

Specifically, we fitted multilevel location-scale meta-regression models with fertiliser type as the sole moderator, and with random effects for study identity and effect-size identity to account for non-independence among comparisons. In these models, the reference level is animal-based fertiliser, so the intercept represents the expected effect (and heterogeneity) for animal-based systems, and the `contrast` coefficient represents the shift for plant-based systems (Fig. 3C–D).

Before turning to the results, we briefly remind the reader that Model 3 refers to a location-scale meta-regression of mean effects (lnRR), whereas Model 4 refers to an analogous location-scale meta-regression of variability (lnCVR).

For lnRR, the location part showed a negative average effect under animal-based fertilisers (Model 3 using `brms`: $\beta_0^{(l)}$ = -0.27, 95% credible interval, CrI = [-0.38, -0.16]), implying lower organic yields on average. The contrast coefficient (animal-plant) is also negative ($\beta^{(l)}_{\texttt{animal-plant}}$ = -0.14, 95% CrI [-0.26, -0.03]), indicating an additional yield penalty under plant-based fertilisers relative to animal-based fertilisers. On the response-ratio scale, these correspond approximately to organic yields of $\exp(-0.27) \approx 0.76$ (about 24% lower than conventional) under animal-based fertilisers, versus $\exp(-0.27 - 0.14) \approx 0.66$ (about 34% lower) under plant-based fertilisers. The scale part (modelled on the log-σ scale) indicated that the plant-based system tended to be more variable (less predictable)($\beta^{(s)}_{\texttt{animal-plant}}$ = 0.17, 95% CrI = [-0.02, 0.37]), corresponding to about a 19% increase in residual heterogeneity for plant-based fertilisers (multiplicative factor $\exp(0.17) \approx 1.19$), with credible overlap around no difference (Fig. 3A,C,E).

For lnCVR, the location part indicated that organic systems tended to have higher mean-corrected variability (lower stability) under animal-based fertilisers (Model 4 using `brms`: $\beta_0^{(l)} = 0.30$, 95% CrI = [0.14, 0.47]; roughly $\exp(0.30) \approx 1.35$, i.e. about 35% higher CV in organic than conventional). The contrast (animal-plant) coefficient is small and highly uncertain ($\beta^{(l)}_{\texttt{animal-plant}}$ =0.08, 95% CrI = [-0.16,0.32]), suggesting that fertiliser type did not strongly shift the average stability difference. In contrast, the scale part shows clear evidence that fertiliser type governs predictability (context dependence) of lnCVR effects ($\beta^{(l)}_{\texttt{animal-plant}}$ =0.52, 95% CrI = [0.18, 0.86]), implying a much wider spread of lnCVR effects across studies for the plant-based fertilisers ($\exp(0.52) \approx 1.68$; $\sigma \approx \exp(-0.83 + 0.52) \approx 0.73$). In other words, while the expected stability gap (lnCVR) is broadly similar between fertiliser categories, the evidence is far less transferable under plant-based fertilisers: some studies show near parity in stability, whereas others show much larger organic-conventional differences (Fig. 3B,D,F).

These paired results illustrate the useful extra insights which location-scale modelling could provide and such insights are potentially decision-relevant. Fertiliser type matters for lnRR largely through the expected yield gap (location), whereas for lnCVR it matters mainly through the consistency of the stability comparison (scale). Thus, a moderator can be important even when it barely shifts the mean effect: it can determine whether an estimated effect is robust and transferable, or whether it is strongly contingent on unmeasured ecological and methodological details within that moderator category (for comparable models using `metafor` and `brms`, see the online tutorial).

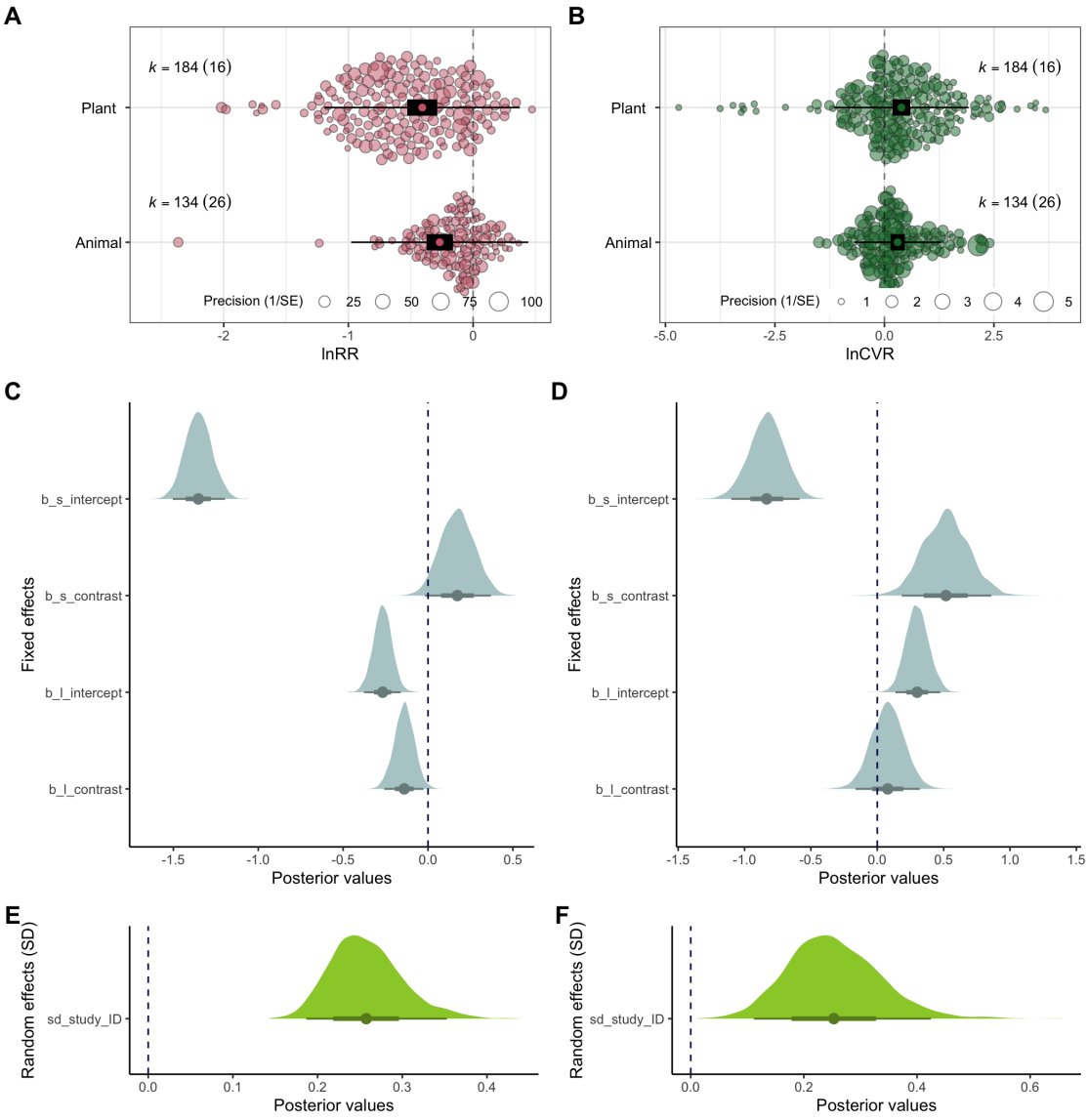

Figure 3: **Location-scale meta-regression showing how a moderator shifts expected effects and their predictability.** Results are shown for lnRR (A,C,E) and lnCVR (B,D,F) when organic fertiliser type (animal- vs plant-derived) is used as a categorical moderator. k denotes the number of effect sizes (number of studies in parentheses). (A,B) Orchard plots stratified by fertiliser type. Points are individual effect sizes (bubble size $\propto$ 1/SE; horizontal jitter for visibility). Black squares and thick horizontal bars show the fitted subgroup mean and its 95% interval; thin whiskers show the 95% prediction interval. The vertical dashed line indicates no difference between organic and conventional farming. (C,D) Posterior distributions for fixed effects from a Bayesian location-scale model: location (mean) parameters ($b_0^{(l)}$ and $b_{\text{contrast(animal}-\text{plant)}}^{(l)}$) and scale (heterogeneity) parameters ($b_0^{(s)}$ and $b_{\text{contrast(animal}-\text{plant)}}^{(s)}$), where the scale coefficients act on $\log(\sigma)$ (positive values indicate greater residual heterogeneity). The intercept corresponds to animal-based fertiliser; the contrast gives the shift for plant-based fertiliser. (E,F) Posterior distributions for the between-study (random-effect) standard deviation. Points are posterior medians; thick and thin intervals denote 66% and 95% credible intervals, respectively. The `R` packages: `ggplot2` (Wickham, 2011), `tidybayes` (Kay, 2020) and `orchaRd` (Nakagawa et al., 2023a) were used to draw these plots.

5 Conclusion and future directions

In this article, we make a simple claim: many ecological syntheses under-use the information already present in standard meta-analytic datasets. When primary studies report group means ($\bar{X}$), standard deviations (s) and sample sizes (n), the same evidence base can usually support inference about (i) average outcomes (via lnRR), (ii) within-population stability or reliability (via lnVR and lnCVR), and (iii) how both of these vary across contexts. The key is to treat "variability" (dispersion within study groups or populations) and "heterogeneity" (dispersion among effect sizes across studies) as distinct targets rather than a single catch-all notion of "variation" (Senior et al., 2016b; Yang et al., 2025). Once that distinction is made, the analytical steps follow naturally: estimate mean effects, estimate variability effects, explain both with moderators, and then test whether predictability itself changes with context using location-scale models (Viechtbauer and López-López, 2022; Nakagawa et al., 2025a).

The practical implication is that moderators can matter in two different ways. Some moderators primarily shift the expected effect (the "location" component), changing what we should expect on average in a new setting. Other moderators primarily shift the predictability of the evidence (the "scale" component), changing how transferable (generalisable) that expectation is (cf., Spake et al., 2022). In the organic-conventional case study (Ponisio et al., 2015), placing lnRR and lnCVR side-by-side highlights why this matters: stakeholders care not only about the expected yield gap, but also about whether organic systems are comparably reliable and whether that reliability comparison is consistent across agronomic contexts (Knapp and van der Heijden, 2018; Nakagawa et al., 2015; Senior et al., 2020). Location–scale modelling formalises that second question, replacing a generic "it depends" with an explicit, testable statement about when effects are more or less heterogeneous (Viechtbauer and López-López, 2022; Nakagawa et al., 2025a).

Three straightforward shifts would make this "one dataset, four analyses" approach routine: one shift for all empiricists and two for meta-analysts. First, primary studies should consistently report the minimal summary statistics that enable both mean and variability synthesis (means, SDs, and sample sizes for each group), alongside design information that induces dependence (shared controls, repeated measures, multi-arm comparisons) (Lajeunesse, 2011; Nakagawa and Santos, 2012). Second, meta-analysts should treat mean-variance relationships as a modelling choice rather than a nuisance: lnVR is informative about absolute spread, whereas lnCVR targets stability relative to mean performance, and reporting both can clarify whether apparent changes in variability are driven mainly by shifts in the mean (Nakagawa et al., 2015; Senior et al., 2020). Third, applied syntheses should present results in prediction-oriented terms: pooled means and pooled variability effects are useful, but decision-relevance often hinges on how widely effects vary across studies and whether moderators identify contexts where results are more transferable (Senior et al., 2016b; Nakagawa et al., 2021).

In short, the contribution here is not a new statistic but a suggested shift in a meta-analytic workflow in ecology:

use the same dataset to estimate performance and stability, and then use the same moderators to explain both the expected effects and their predictability (Nakagawa et al., 2015; Senior et al., 2020; Viechtbauer and López-López, 2022; Nakagawa et al., 2025a). That workflow better matches what ecologists and decision-makers typically want to know, not only "what happens on average," but "how consistent is it," and "how generalisable is this result across different contexts?" (Senior et al., 2016b; Yang et al., 2025; Spake et al., 2023). With this new workflow, let us make full use of the four meta-analytic models to better understand our changing world.

Data Availability Statement

All data, scripts and relevant files used for this study can be found at the GitHub repository (https://anonymous.4open.science/r/one-dataset-four-meta-analyses-3B43/) and a version of it will be archived at Zenodo (link).

References

M. E. Brooks, K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2):378–400, 2017. doi: 10.32614/RJ-2017-066.

S. P. Brugger, I. Angelescu, A. Abi-Dargham, R. Mizrahi, V. Shahrezaei, and O. D. Howes. Heterogeneity of striatal dopamine function in schizophrenia: meta-analysis of variance. *Biological psychiatry*, 87(3):215–224, 2020.

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

S. A. Chamberlain, J. L. Bronstein, and J. A. Rudgers. How context dependent are species interactions? *Ecology letters*, 17(7):881–890, 2014.

I. R. Cleasby, S. Nakagawa, and H. Schielzeth. 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, 2015.

R. P. Duncan and B. J. Kefford. Interactions in statistical models: three things to know. *Methods in Ecology and Evolution*, 12(12):2287–2297, 2021.

L. V. Hedges, J. Gurevitch, and P. S. Curtis. The Meta-Analysis of Response Ratios in Experimental Ecology. *Ecology*, 80(4):1150–1156, 1999. ISSN 1939-9170. doi: 10.1890/0012-9658(1999)080[1150:TMAORR]2.0.CO;2. URL https://onlinelibrary.wiley.com/doi/abs/10.1890/0012-9658%281999%29080%5B1150%3ATMAORR%5D2.0.CO%3B2. _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/0012-9658%281999%29080%5B1150%3ATMAORR%5D2.0.CO%3B2.

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

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

W. G. Hill and H. A. Mulder. Genetic analysis of environmental variation. *Genetics Research*, 92(5-6):381–395, 2010.

O. D. Howes and G. E. Chapman. Understanding variability: the role of meta-analysis of variance. *Psychological Medicine*, 54(12):3233–3236, Sept. 2024. ISSN 0033-2917, 1469-8978. doi: 10.1017/S0033291724001971. URL https://www.cambridge.org/core/journals/psychological-medicine/article/understanding-variability-the-role-of-metaanalysis-of-variance/3628110CD02FAFB7ECBF5C77AC6B21E5.

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

S. Knapp and M. G. A. van der Heijden. A global meta-analysis of yield stability in organic and conservation agriculture. *Nature Communications*, 9(1):3632, Sept. 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05956-1. URL https://www.nature.com/articles/s41467-018-05956-1. Publisher: Nature Publishing Group.

K. Kristensen, M. McGillycuddy, and C. Williams. Covariance structures with glmmtmb. Vignette for the R package `glmmTMB` on the Comprehensive R Archive Network (CRAN), Jan. 2026. URL https://cran.r-project.org/web/packages/glmmTMB/vignettes/covstruct.html%7D. Accessed: 2026-01-28.

M. J. Lajeunesse. Achieving synthesis with meta-analysis by combining and comparing all available studies. *Ecology*, 91(9):2561–2564, Sept. 2010. ISSN 0012-9658, 1939-9170. doi: 10.1890/09-1530.1. URL https://esajournals.onlinelibrary.wiley.com/doi/10.1890/09-1530.1.

M. J. Lajeunesse. On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology*, 92(11):2049–2055, 2011. ISSN 1939-9170. doi: 10.1890/11-0423.1. URL https://onlinelibrary.wiley.com/doi/abs/10.1890/11-0423.1. _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/11-0423.1.

M. J. Lajeunesse. Bias and correction for the log response ratio in ecological meta-analysis. *Ecology*, 96(8):2056–2063, 2015.

J. H. Lawton. Are there general laws in ecology? *Oikos*, pages 177–192, 1999.

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

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

S. A. Levin. The problem of pattern and scale in ecology: the robert h. macarthur award lecture. *Ecology*, 73(6): 1943–1967, 1992.

H. A. Mulder, P. Bijma, and W. G. Hill. Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genetics Selection Evolution*, 40:1–23, 2008.

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

S. Nakagawa and E. S. A. Santos. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, 26(5):1253–1274, Sept. 2012. ISSN 1573-8477. doi: 10.1007/s10682-012-9555-5. URL https://doi.org/10.1007/s10682-012-9555-5.

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

2015. ISSN 2041-210X. doi: 10.1111/2041-210X.12309. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12309. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12309.

S. Nakagawa, M. Lagisz, R. E. O'Dea, J. Rutkowska, Y. Yang, D. W. A. Noble, and A. M. Senior. The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. *Research Synthesis Methods*, 12(1):4–12, 2021. ISSN 1759-2887. doi: 10.1002/jrsm.1424. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1424. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.1424.

S. Nakagawa, M. Lagisz, R. E. O'Dea, P. Pottier, J. Rutkowska, A. M. Senior, Y. Yang, and D. W. A. Noble. orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. *Methods in Ecology and Evolution*, 14 (8):2003–2010, 2023a. ISSN 2041-210X. doi: 10.1111/2041-210X.14152. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14152. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.14152.

S. Nakagawa, D. W. Noble, M. Lagisz, R. Spake, W. Viechtbauer, and A. M. Senior. A robust and readily implementable method for the meta-analysis of response ratios with and without missing standard deviations. *Ecology Letters*, 26(2):232–244, 2023b.

S. Nakagawa, A. Mizuno, K. Morrison, L. Ricolfi, C. Williams, S. M. Drobniak, M. Lagisz, and Y. Yang. Location-Scale Meta-Analysis and Meta-Regression as a Tool to Capture Large-Scale Changes in Biological and Methodological Heterogeneity: A Spotlight on Heteroscedasticity. *Global Change Biology*, 31(5):e70204, 2025a. ISSN 1365-2486. doi: 10.1111/gcb.70204. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.70204. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.70204.

S. Nakagawa, A. Mizuno, C. Williams, M. Lagisz, Y. Yang, and S. M. Drobniak. Quantifying macro-evolutionary patterns of trait mean and variance with phylogenetic location–scale models. *Methods in Ecology and Evolution*, 16(11):2585–2602, 2025b.

S. Nakagawa, S. Ortega, E. Gazzea, M. Lagisz, A. Lenz, E. Lundgren, and A. Mizuno. Location–scale models in ecology and evolution: Heteroscedasticity in continuous, count and proportion data. *Methods in Ecology and Evolution*, 2025c.

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

R. E. O'Dea, M. Lagisz, M. D. Jennions, and S. Nakagawa. Gender differences in individual variation in academic grades fail to fit expected patterns for stem. *Nature communications*, 9(1):3777, 2018.

S. D. Peacor, C. Song, J. R. Bence, A. A. Briggs, E. A. Hamman, and C. W. Osenberg. Ecological meta-analyses often produce unwarranted results. *Ecology*, 106(12):e70269, 2025.

L. C. Ponisio, L. K. M'Gonigle, K. C. Mace, J. Palomino, P. de Valpine, and C. Kremen. Diversification practices reduce organic to conventional yield gap. *Proceedings. Biological Sciences*, 282(1799):20141396, Jan. 2015. ISSN 1471-2954. doi: 10.1098/rspb.2014.1396.

R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54(3):507–554, 2005.

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

A. M. Senior, A. K. Gosby, J. Lu, S. J. Simpson, and D. Raubenheimer. Meta-analysis of variance: an illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, Medicine, and Public Health*, 2016(1):244–255, Jan. 2016a. ISSN 2050-6201. doi: 10.1093/emph/eow020. URL https://doi.org/10.1093/emph/eow020.

A. M. Senior, C. E. Grueber, T. Kamiya, M. Lagisz, K. O'Dwyer, E. S. A. Santos, and S. Nakagawa. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology*, 97(12):3293–3299, 2016b. ISSN 1939-9170. doi: 10.1002/ecy.1591. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.1591. _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.1591.

A. M. Senior, W. Viechtbauer, and S. Nakagawa. Revisiting and expanding the meta-analysis of variation: The log coefficient of variation ratio. *Research Synthesis Methods*, 11(4):553–567, 2020. ISSN 1759-2887. doi: 10.1002/jrsm.1423. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1423. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.1423.

R. Spake, R. E. O'dea, S. Nakagawa, C. P. Doncaster, M. Ryo, C. T. Callaghan, and J. M. Bullock. Improving quantitative synthesis to achieve generality in ecology. *Nature Ecology & Evolution*, 6(12):1818–1828, 2022.

R. Spake, D. E. Bowler, C. T. Callaghan, S. A. Blowes, C. P. Doncaster, L. H. Antao, S. Nakagawa, R. McElreath, and J. M. Chase. Understanding 'it depends' in ecology: a guide to hypothesising, visualising and interpreting statistical interactions. *Biological Reviews*, 98(4):983–1002, 2023.

D. M. Stasinopoulos and R. A. Rigby. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23:1–46, 2008.

A. Sánchez-Tójar, N. P. Moran, R. E. O'Dea, K. Reinhold, and S. Nakagawa. Illustrating the importance of meta-analysing variances alongside means in ecology and evolution. *Journal of Evolutionary Biology*, 33(9):1216–1223, Sept. 2020. ISSN 1010-061X. doi: 10.1111/jeb.13661. URL https://doi.org/10.1111/jeb.13661.

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

W. Viechtbauer and J. A. López-López. Location-scale models for meta-analysis. *Research Synthesis Methods*, 13(6): 697–715, 2022. ISSN 1759-2887. doi: 10.1002/jrsm.1562. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1562. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jrsm.1562.

D. F. Westneat, M. Schofield, and J. Wright. Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, 24(3):598–604, 2013.

H. Wickham. ggplot2. *Wiley interdisciplinary reviews: computational statistics*, 3(2):180–185, 2011.

C. Williams, Y. Yang, D. I. Warton, and S. Nakagawa. Modelling approaches for meta-analyses with dependent effect sizes in ecology and evolution: A simulation study. *Methods in Ecology and Evolution*, 16(10):2362–2379, 2025.

Y. Yang, D. W. Noble, R. Spake, A. M. Senior, M. Lagisz, and S. Nakagawa. A pluralistic framework for measuring, interpreting and decomposing heterogeneity in meta-analysis. *Methods in Ecology and Evolution*, 16(11):2710–2725, 2025.