xxx xxx (0000), **xx**:xx 1–**??**

Tutorial

A unified framework for phylogenetic and spatial meta-analysis: concepts, implementation, and practical guidance

Ayumi Mizuno[1], Coralie Williams[2,3], Malgorzata Lagisz[1,2], Alistair M. Senior[4] and Shinichi Nakagawa[1,2]

[1]Collaboration for Open Science and Synthesis in Ecology and Evolution (COSSEE), Department of Biological Sciences, University of Alberta, Edmonton, T6G 2E9, Alberta, Canada.
[2]Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, 2052, New South Wales, Australia.
[3]School of Mathematics and Statistics, University of New South Wales, Sydney, 2052, New South Wales, Australia.
[4]Charles Perkins Centre (CPC) and Sydney Precision Data Science Centre (SPDSC), School of Life and Environmental Sciences, Faculty of Science, University of Sydney, Sydney, 2006, New South Wales, Australia.

Keywords: Hierarchical modelling, Quantitative analysis, Brownian motion, Ornstein–Uhlenbeck process

Abstract

Meta-analyses in ecology and evolution, and related fields, can include effect sizes structured by shared evolutionary history or spatial distance. In this tutorial paper, we show that phylogenetic and spatial meta-analyses can be formulated within the same theoretical framework based on correlated random effects. From this perspective, the two approaches differ only in how distance is defined: evolutionary time in phylogenetic meta-analyses versus geographic distance in spatial ones, while sharing the same underlying statistical logic. This unified view clarifies relationships among commonly used correlation structures and reveals their direct correspondence across phylogenetic and spatial settings. Building on this framework, we illustrate how researchers can implement phylogenetic and spatial meta-analytic models in several widely used `R` packages, including `metafor`, `glmmTMB`, and `brms`. Using published datasets, we demonstrate how researchers can express equivalent model specifications across frequentist and Bayesian frameworks and how these models allocate variance across hierarchical levels. We also present practical issues related to model identifiability and data structure and highlight considerations for specifying and interpreting correlated meta-analytic models. Although we draw examples from ecology and evolutionary biology, the same framework can be applied to meta-analyses in many other fields, for example, epidemiology and public health, education and social policy, and linguistics and cultural evolution.

Highlights

What is already known

- Meta-analytic datasets in ecology and evolution often show phylogenetic and spatial non-independence.
- These two types of non-independence share key conceptual and statistical bases, but are often considered separately.

What is new

- We provide a first integrated tutorial on phylogenetic and spatial meta-analysis.
- We highlight both the similarities and the differences between these approaches in their underlying concepts, data preparation, and model fitting.

Potential impact for *Research Synthesis Methods* readers

- This tutorial helps reduce the conceptual divide between phylogenetic and spatial meta-analytic modelling.
- It encourages readers to focus on data structure, identifiability, and model choice when handling non-independence in meta-analytic datasets.

1. Introduction

Meta-analyses in ecology and evolution are often complicated by multiple sources of non-independence, including shared evolutionary history and spatial proximity[1,2]. Such dependencies violate the assumption that effect sizes are statistically independent and, if ignored, can bias estimates and lead to overconfident conclusions[2,3]. Dealing with non-independence is a key challenge in ecological and evolutionary synthesis.

A major source of non-independence is phylogeny, as effect sizes from closely related species are often more similar than expected under independence. Phylogenetic meta-analysis handles this by incorporating evolutionary relatedness into the model, allowing effect sizes to covary according to shared ancestry rather than treating them as independent observations[1,4–9].

When primary studies are drawn from different geographic regions, effect sizes from nearby locations are usually more similar than those from distant places, leading to spatial non-independence[10]. Spatial meta-analytic models can address this by incorporating correlations as a function of geographic distance[11]. Conceptually, this is analogous to phylogenetic meta-analysis: both extend meta-analytic models by allowing correlations among effect sizes to depend on distance, whether geographic or evolutionary (Figure 1).

Despite this conceptual similarity, phylogenetic and spatial meta-analyses differ markedly in their practical barriers to implementation. Phylogenetic meta-analysis can be relatively straightforward once a phylogeny is available. However, models are not always specified to include both phylogenetic and non-phylogenetic variance components, potentially leading to inadequate representations of evolutionary similarity among species[5,9]. Spatial meta-analysis requires additional data and modelling decisions, such as sampling coordinates and distance-based correlation functions. This creates additional opportunities for model misspecification, with consequences for bias and uncertainty that parallel those in phylogenetic meta-analysis.

Here, we present a unified framework for phylogenetic meta-analysis and show how it extends naturally to spatial meta-analysis. Although these approaches are often discussed separately, they can be understood within a shared conceptual and modelling workflow. To complement this paper, we provide an online tutorial demonstrating how to implement both phylogenetic and spatial meta-analyses using frequentist and Bayesian approaches in the `R` packages `metafor`[12], `glmmTMB`[13], and `brms`[14]. The online tutorial also emphasises model specification, interpretation, and common problems (https://ayumi-495.github.io/phylo_spatial_tutorial/).

2. Models

We begin with a basic random-effects meta-analytic model that accounts for among-study variability. We then show how this model can be extended to accommodate phylogenetic and spatial non-independence. Fixed-effect models are not considered further because they assume a common true effect size across all studies and ignore among-study heterogeneity. These assumptions are rarely suitable in ecological and evolutionary research[8,15].

2.1. Basic random-effects model (among-study heterogeneity)

Random-effects models are used in meta-analysis to accommodate heterogeneity among studies. In its simplest form, the model includes a study-level random effect:

$$\begin{aligned} y_j &= \beta_0 + u_j + e_j, \\ u_j &\sim \mathcal{N}(0, \sigma_u^2), \\ e_j &\sim \mathcal{N}(0, v_j). \end{aligned} \tag{1}$$

Here, u_j is the random effect for the j-th study, representing study-specific deviations from the overall mean, and σ_u^2 quantifies among-study heterogeneity. The sampling error term e_j has known variance v_j. In conventional meta-analytic notation, σ_u^2 is often denoted by τ^2. We use σ_u^2 throughout this paper, but τ^2 is also used in `metafor` output and for the variance parameter associated with distance-based correlation structures. When $\sigma_u^2 = 0$, the model reduces to a fixed-effect model in which all variation is attributed to sampling error alone. This random-effects model provides the starting point for the extensions considered below.

2.2. Multilevel model (effect sizes nested within studies)

While the random-effects model in eq.1 accounts for variation among studies, it assumes that each study contributes only a single effect size. This assumption is often violated in ecological and evolutionary meta-analyses, where multiple effect sizes are frequently reported from the same study[3,16]. In such cases, effect sizes from the same study cannot be treated as fully independent. A multilevel (hierarchical) model extends the random-effects model by introducing an additional level of random variation to reflect this nested structure.

$$\begin{aligned} y_i &= \beta_0 + u_{j[i]} + s_i + e_i, \\ u_j &\sim \mathcal{N}(0, \sigma_u^2), \\ s_i &\sim \mathcal{N}(0, \sigma_s^2), \\ e_i &\sim \mathcal{N}(0, v_i), \end{aligned} \tag{2}$$

where y_i denotes the i-th observed effect size, and $j[i]$ indicates the study from which that effect size i was derived. The term u_j is a study-level random effect with variance σ_u^2, capturing among-study heterogeneity in the study-average effect size. The term s_i is an observation-level random effect with variance σ_s^2, capturing residual heterogeneity among effect sizes within the same study that is not explained by sampling error. Together, these components imply that true effect sizes vary randomly around the overall mean β_0, both among and within studies.

This model explicitly accounts for the fact that effect sizes derived from the same study are not statistically independent. Such dependencies can arise from overlapping samples, repeated measurements on the same individuals, or other shared sources of variation[17,18]. In these cases, effect sizes share common variance components and cannot be treated as fully independent observations[19]. By modelling both among- and within-study heterogeneity (via σ_u^2 and σ_s^2, respectively), the multilevel model provides a more realistic and statistically appropriate representation of the hierarchical structure of meta-analytic data[20].

2.3. Phylogenetic multilevel meta-analysis

Meta-analyses in evolutionary ecology often synthesise effect sizes across multiple species to address broad questions about ecological responses and trait evolution. Because closely related species tend to resemble one another due to shared evolutionary history[21,22], effect sizes drawn from different species may not be statistically independent. Ignoring this phylogenetic non-independence can underestimate uncertainty and inflate type-I error rates, particularly when many effect sizes come from closely related species[9]. To account for this, the multilevel meta-analytic model in eq. 2 can be extended to include species-level random effects that partition variation into phylogenetically structured and non-phylogenetic components[4,5,9,23].

$$
\begin{aligned}
y_i &= \beta_0 + u_{j[i]} + p_{k[i]} + q_{k[i]} + s_i + e_i, \\
u_j &\sim \mathcal{N}(0, \sigma_u^2), \\
p_k &\sim \mathcal{N}(0, \sigma_p^2 \mathbf{A}), \\
q_k &\sim \mathcal{N}(0, \sigma_q^2), \\
s_i &\sim \mathcal{N}(0, \sigma_s^2), \\
e_i &\sim \mathcal{N}(0, v_i),
\end{aligned}
\tag{3}
$$

where $p_{k[i]}$ is the phylogenetic random effect associated with the k-th species to which effect size i belongs. These effects follow a multivariate normal distribution with covariance matrix $\sigma_p^2\mathbf{A}$, where $\mathbf{A}$ is the phylogenetic correlation matrix derived from a species tree. This structure implies that closely related species are expected to have more similar true effect sizes than distantly related species. The model also includes a non-phylogenetic species-level random effect, $q_{k[i]} \sim \mathcal{N}(0, \sigma_q^2)$, which captures additional among-species heterogeneity not explained by shared ancestry. Including both $p_{k[i]}$ and $q_{k[i]}$ allows species-level variation to be decomposed into phylogenetically structured and non-phylogenetic components, which is central to interpreting heterogeneity in cross-species meta-analysis[23].

Under this model, the species-level deviation from the overall mean is given by $p_k + q_k$. The intercept β_0 represents a global mean across the analysed species and studies, after accounting for sampling variance and hierarchical heterogeneity. Because the phylogenetic effect p_k is modelled as a mean-zero deviation structured by $\mathbf{A}$, β_0 should not be interpreted as the trait value or effect size of a common ancestor. Such an interpretation would require an alternative parametrisation in which the root state is modelled explicitly. Simulation work suggests that omitting $q_{k[i]}$ can distort variance partitioning, often inflating estimates of phylogenetic variance when non-phylogenetic species-level variation is present, whereas including $q_{k[i]}$ has little cost when such variation is absent[9]. In light of these findings, unless there is exactly one effect size measurement per species or a strong justification, both $p_{k[i]}$ and $q_{k[i]}$ should be retained.

A further practical issue is model identifiability, that is, whether the data contain enough information to estimate the model parameters reliably. Poor identifiability can lead to unstable estimates, non-convergence, or misleading variance partitioning[1,9] (**Box 1**). In this setting, separating phylogenetic and non-phylogenetic species-level variation is most feasible when species contribute multiple effect sizes and when study and species identities are not strongly confounded. Identifiability problems are most severe when several hierarchical levels are included, but replication is limited. We recommend evaluating sample sizes at each level, including the numbers of studies and species and the distribution of effect sizes per species, before fitting the full model.

2.4. Spatial meta-analysis model

In ecological and environmental studies, effect sizes can be associated with specific geographic locations, and nearby locations may yield more similar values than those farther apart, a phenomenon known as spatial autocorrelation[10]. To account for this dependence, we build on the multilevel model in Section 2.2 (eq. 2) by allowing the study-level random effect to follow a distance-based spatial correlation structure. As a conceptual starting point, a basic spatial meta-analytic model can be written as:

$$
\begin{aligned}
y_i &= \beta_0 + l_{j[i]} + s_i + e_i, \\
l_j &\sim \mathcal{N}(0, \sigma_l^2 \mathbf{M}), \\
s_i &\sim \mathcal{N}(0, \sigma_s^2), \\
e_i &\sim \mathcal{N}(0, v_i),
\end{aligned}
\tag{4}
$$

where y_i is the i-th observed effect size, β_0 is the overall mean effect size, and $j[i]$ denotes the study associated with effect size i. Here, we assume that each study has exactly one sampling location with known coordinates, so all effect sizes within a study share the same location. The term $l_{j[i]}$ is a spatially structured random effect with covariance matrix $\sigma_l^2\mathbf{M}$, where $\mathbf{M}$ is a spatial correlation matrix derived from pairwise geographic distances among study locations. Thus, for locations x and y, $\mathrm{Cov}(l_x, l_y) = \sigma_l^2 M_{xy}$. The term s_i captures residual within-study heterogeneity, and e_i denotes sampling error with known variance v_i. Note that σ_l^2 is reported as τ^2 in `metafor`'s distance-based correlation structures, which can be confused with the conventional use of τ^2 for among-study heterogeneity in standard random-effects meta-analysis.

More complex spatial models are possible. For example, one may wish to distinguish study-level variation from location-level variation, allow for multiple studies at the same location, or separate spatially structured from unstructured location-level heterogeneity. A conceptual extension is:

$$\begin{aligned} y_i &= \beta_0 + u_{j[i]} + l_{h[i]} + m_{h[i]} + s_i + e_i, \\ u_j &\sim \mathcal{N}(0, \sigma_u^2), \\ l_h &\sim \mathcal{N}(0, \sigma_l^2\mathbf{M}), \\ m_h &\sim \mathcal{N}(0, \sigma_m^2), \\ s_i &\sim \mathcal{N}(0, \sigma_s^2), \\ e_i &\sim \mathcal{N}(0, v_i), \end{aligned} \tag{5}$$

where $u_{j[i]}$ is a study-level random effect, $h[i]$ denotes the geographic location of effect size i, $l_{h[i]}$ and $m_{h[i]}$ represent spatially structured and unstructured location-level random effects, respectively. In practice, however, models such as eq. 5 are often weakly identifiable in meta-analytic datasets. Distinguishing spatially structured from unstructured location-level variation requires strong and well-balanced replication across locations, which is rarely available. Identifiability may be further reduced when geographic coordinates are missing, imprecise, unevenly distributed, or concentrated at only very short or very long distances. For these reasons, eq. 4 usually provides a more parsimonious and practically estimable starting point, whereas more complex formulations should be treated cautiously and assessed using sensitivity analyses (see **Box 1**).

2.5. Correlation

Phylogenetic and spatial correlation structures share a common mathematical basis: both model correlation as a decreasing function of distance. The key difference lies in what distance represents (Figure 1). In phylogenetic models, t_{xy} denotes the evolutionary distance between species x and y, typically measured as the sum of branch lengths connecting them in a phylogenetic tree. In spatial models, d_{xy} denotes the geographic distance between locations x and y. In both cases, the correlation-distance relationship is specified by a kernel function. Different kernels imply different assumptions about how quickly similarity decays with distance. This shared formulation allows phylogenetic and spatial correlation structures to be compared directly. Below, we compare three representative structures and their phylogenetic-spatial counterparts (Figure 2): (1) the linear kernel in spatial models and Brownian motion in phylogenetic models; (2) the exponential kernel and Ornstein-Uhlenbeck; and (3) the squared exponential (Gaussian), which is used only in spatial models.

2.5.1. Why compare correlation structures?

In reality, the appropriate correlation structure is hard to know with certainty and should be treated as a modelling assumption. Alternative structures are worth considering for three complementary reasons. First, a particular structure may be favoured *a priori* based on theory or quantitative knowledge of

the underlying process, for example, whether similarity is expected to decay gradually or rapidly with distance (Figure 2). Second, fitting alternative structures provides a sensitivity analysis of whether key inferences, such as the overall mean effect or variance components, are robust to this assumption. Third, models can be compared using fit or predictive criteria, such as information criteria or cross-validation, and in some cases, the scientific question itself may centre on comparing competing dependence models. A practical strategy is to begin with commonly used structures, such as Brownian motion or Ornstein-Uhlenbeck for phylogenetic models, and exponential or squared-exponential kernels for spatial models. One or two plausible alternatives can then be fitted to assess whether the substantive conclusions depend on the assumed correlation structure.

2.5.2. Brownian Motion (BM) model / Linear kernel

The Brownian motion (BM) model assumes that traits evolve as a neutral random walk along the branches of a phylogeny[21,24]. Under BM, the covariance between two species is proportional to their shared evolutionary history, quantified by the shared branch length from the root to their most recent common ancestor, $t_{\text{shared}(x,y)}$. Standardising this covariance yields the phylogenetic correlation:

$$C_{\text{BM}}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} = \frac{\sigma^2 t_{\text{shared}(x,y)}}{\sqrt{\sigma^2 t_x \cdot \sigma^2 t_y}} = \frac{t_{\text{shared}(x,y)}}{\sqrt{t_x \cdot t_y}}. \tag{6}$$

For ultrametric trees, where all species are equidistant from the root ($t_x = t_y = T$), eq. 6 simplifies to $C_{\text{BM}}(x, y) = 1 - t_{\text{div}}/T$. Under this special case, phylogenetic correlation declines linearly with divergence time t_{div} (Figure 2). More generally, BM implies that correlation decreases with evolutionary separation, but does not reach exactly zero because all species share some ancestry through the root. For non-ultrametric trees, the relationship is not exactly linear in this simple form, but correlation still declines with evolutionary separation.

In spatial analyses, the corresponding linear kernel assumes that correlation decreases linearly with geographic distance d_{xy} up to a maximum range ρ:

$$C_{\text{lin}}(d_{xy}) = \begin{cases} 1 - \frac{d_{xy}}{\rho}, & \text{if } d_{xy} < \rho, \\ 0, & \text{if } d_{xy} \geq \rho. \end{cases} \tag{7}$$

The BM model is analogous to a linear spatial kernel, in that both imply a linear decline in correlation with distance. The main difference is that the spatial kernel typically includes a finite cutoff, whereas under BM phylogenetic correlation persists, albeit weakly, even among distantly related species.

2.5.3. Ornstein-Uhlenbeck (OU) model / Exponential kernel

The Ornstein-Uhlenbeck (OU) model extends the Brownian motion framework by incorporating attraction towards an optimal value over time, often interpreted as stabilising selection[25,26]. Under the OU model, phylogenetic correlation decays exponentially with evolutionary separation:

$$C_{\text{OU}}(t_{xy}) = \exp(-\alpha t_{xy}), \tag{8}$$

where t_{xy} denotes the evolutionary separation between species x and y, and $\alpha > 0$ controls the rate of decay. Here we present the correlation form; in the full OU process, the covariance is additionally scaled by σ^2. Larger values of α imply more rapid decay, so that correlations among distantly related species become negligible more quickly than under Brownian motion. Unlike BM, which implies linear decline under ultrametric trees, the OU model produces a steep initial decline that then flattens asymptotically while remaining positive for any finite t_{xy} (Figure 2).

In spatial statistics, the corresponding exponential kernel is:

$$C_{\text{exp}}(d_{xy}) = \exp(-d_{xy}/\rho), \tag{9}$$

where d_{xy} is the geographic distance between locations x and y, and $\rho > 0$ is a range parameter controlling the rate of decay. Larger values of ρ imply slower spatial decay. The OU model and the exponential kernel are mathematically equivalent when $\alpha = 1/\rho$. They differ only in the definition of distance, evolutionary in one case and geographic in the other, and in the interpretation of the decay parameter, but they generate correlation matrices in the same way (Figures 1 and 2).

2.5.4. Squared exponential, Gaussian, kernel (spatial only)

The squared exponential, or Gaussian, kernel is widely used in spatial Gaussian process modelling and assumes that correlation decreases with the squared distance between two locations:

$$C_{\text{gau}}(d_{xy}) = \exp\left(-\frac{d_{xy}^2}{\rho^2}\right), \tag{10}$$

where d_{xy} is the geographic distance and $\rho > 0$ is a scale parameter controlling the rate of decay. Compared with the exponential kernel, the squared exponential implies a smoother spatial process and a faster decline in correlation at larger distances. For example, when $d_{xy} = 2\rho$, the correlation is $\exp(-4) \approx 0.018$. Because this kernel implies an extremely smooth spatial process, nearby locations are expected to have highly similar random effects, and the fitted spatial surface changes gradually across space, with abrupt local jumps being unlikely. Unlike the linear and exponential forms discussed above, the squared exponential kernel has no commonly used phylogenetic counterpart. In practice, it is mainly used in spatial models, and it illustrates that spatial analyses often allow a broader range of kernel choices than phylogenetic models (Figure 2).

3. Preparation for analysis

3.1. Phylogenetic matrices

When a phylogeny with branch lengths is available, constructing a phylogenetic correlation matrix is relatively straightforward. For several well-studied taxa, such as birds[27,28], mammals[29], and amphibians[30], such phylogenies are publicly available. Commonly used `R` packages, including `ape` and `phytools`, can be used to compute variance-covariance or correlation matrices under Brownian motion, where the expected correlation structure follows directly from the tree. Alternative evolutionary models, such as Ornstein-Uhlenbeck or Pagel's λ, require additional parameters to be specified or estimated before matrix construction.

In practice, however, phylogenetic information is often incomplete or only partially resolved, particularly in large-scale or multi-taxon meta-analyses. Before constructing a correlation matrix, we need to preprocess the datasets, including assigning branch lengths to topological trees, resolving polytomies, and reconciling species names between the phylogeny and the dataset. These steps are important because inconsistencies at this stage can distort the resulting correlation matrix and downstream meta-analytic inference. A concise summary of common issues and suggested solutions is provided in **Box 2**. Further implementation details and worked examples are available in the accompanying online tutorial.

3.2. Spatial matrices

A spatial correlation matrix reflects the assumption that geographically proximate observations tend to be more similar than those farther apart. Its construction typically involves three steps: obtaining spatial

coordinates, calculating pairwise distances, and converting those distances into correlations using a decreasing kernel function.

Spatial information is usually represented as geographic coordinates, such as latitude and longitude, corresponding to sampling sites, observation points, or derived summaries such as range centroids. These data may come from field GPS records, biodiversity databases such as GBIF or eBird, published sources, or spatial layers such as shapefiles or rasters. Once assembled, coordinates should be stored in a consistent format and projection.

Pairwise distances among locations can then be calculated using functions such as `distm()` from `geosphere` or `st_distance()` from `sf`, depending on whether spherical or projected distances are required. A spatial correlation matrix is obtained by applying a monotonic decreasing kernel, such as the exponential or squared-exponential form, to the resulting distance matrix. When constructing such matrices manually, it is important to ensure consistency in both distance units and geometry. Mixing, for example, degrees and kilometres or spherical and Euclidean distances can lead to misleading correlation structures.

Some modelling packages, including `metafor` and `brms`, also allow users to supply coordinates directly and apply the chosen spatial correlation structure internally during model fitting. Manually creating distance or correlation matrices is generally unnecessary unless a custom structure is needed. Further implementation details and worked examples are provided in the accompanying tutorial using `metafor`, `glmmTMB`, and `brms`.

4. Illustrative examples

To demonstrate the implementation of the models introduced above, we present a series of worked examples using publicly available datasets. In each example, we fit an intercept-only meta-analytic model to estimate the overall mean effect size, assuming independent sampling errors for simplicity. Models are fitted using three widely used `R` packages: `metafor`[12], `glmmTMB`[13], and `brms`[14], and we compare the resulting estimates and variance components across packages.

In practice, each analysis follows the same broad workflow: preparing the meta-analytic dataset, identifying the units over which non-independence is expected, constructing the relevant dependence structure, and fitting the corresponding meta-analytic model. For phylogenetic analyses, this involves matching species to a phylogeny and constructing a phylogenetic correlation matrix. For spatial analyses, it involves assembling sampling coordinates, calculating pairwise distances, and converting them into a spatial correlation matrix using an appropriate kernel function.

We also visualise the `metafor` outputs using `orchaRd`[31], provide guidance on reporting, and include pseudocode to highlight the hierarchical structure of each model and the role of its random effects. Full code and reproducible examples are available in the online tutorial.

4.1. Phylogenetic meta-analysis

Using the dataset of Moura et al.[32], available in the `metadat` `R` package[33], we illustrate the phylogenetic multilevel meta-analytic model introduced in eq. 3. We analyse the data under a Brownian motion (BM) model as the default phylogenetic correlation structure, and briefly discuss how an alternative structure, the Ornstein-Uhlenbeck (OU) model, can also be implemented.

The dataset contains 1,828 effect sizes from 457 studies across 341 animal species spanning multiple taxa. The effect sizes quantify size-assortative mating, that is, the strength and direction of the association between male and female body size within mating pairs. Most were originally reported as Pearson's correlation coefficients and analysed after Fisher's Z_r transformation, with other statistics converted to Z_r where necessary.

In all three packages, `metafor`, `glmmTMB`, and `brms`, we specified the same random-effects structure corresponding to eq. 3: study identity as a among-study random effect, effect-size identity as a within-study random effect, species identity as a non-phylogenetic species-level random effect, and a phylogenetic random effect based on the correlation matrix **A**.

Matrix **A** was derived from an Open Tree of Life Metazoa phylogeny pruned to the species represented in the dataset. Because the resulting tree did not contain branch lengths in the required form, branch lengths were assigned using `compute.brlen()` before constructing the phylogenetic correlation matrix.

For this example, the model can be written as:

$$
\begin{aligned}
y_i &= \beta_0 + \text{Study\_id}_{j[i]} + \text{Effect\_id}_i \\
&\quad + \text{Species\_phylo}_{k[i]} + \text{Species\_nonphylo}_{k[i]} + \text{Sampling error}_i, \\
\text{Study\_id}_j &\sim \mathcal{N}(0, \sigma_u^2), \quad \text{Effect\_id}_i \sim \mathcal{N}(0, \sigma_s^2) \\
\text{Species\_phylo}_k &\sim \mathcal{N}(0, \sigma_p^2 \mathbf{A}), \quad \text{Species\_nonphylo}_k \sim \mathcal{N}(0, \sigma_q^2) \\
\text{Sampling error}_i &\sim \mathcal{N}(0, v_i).
\end{aligned}
$$

The three R packages differ in how the phylogenetic structure is specified: In `metafor`, the correlation matrix **A** is supplied via the R argument to the `rma.mv` function.

```
# yi: effect sizes;
# vi: known sampling variances
# Study_id: unique study identifier for the among-study random effect
# Effect_id: unique effect-size identifier for the within-study random
    effect
# Species_phylo: species names used for the phylogenetic random effect (
    linked to A)
# Species_nonphylo: the non-phylogenetic species effect
# A: phylogenetic correlation matrix with row/column names matching levels
    (Species_phylo)

A <- vcv.phylo(tree, corr = TRUE)

fit_metafor <- rma.mv(yi, vi,
              random = list(~1|Study_id,
                            ~1|Effect_id,
                            ~1|Species_phylo,
                            ~1|Species_nonphylo),
              R = list(Species_phylo = A),
              data = data)
```

We can also fit phylogenetic meta-analysis models in `glmmTMB` using `propto()` and the novel `equalto()` covariance structures (glmmTMB vignette and Williams et al[34,35]). The phylogenetic covariance matrix is passed to `propto()`, and the known sampling variances are passed to `equalto()`. Here, `VCV` is the diagonal matrix of sampling variances `vi`, and `g` is a grouping factor required by both `equalto()` and `propto()`. The same model can also be written by adding an observation-level random effect, `(1|id)`, and fixing the residual variance to zero with `dispformula = 0`, which makes the sampling-variance component explicit.

```
A <- A[sort(rownames(A)), sort(rownames(A))]

# glmmTMB requires a grouping factor for equalto() / propto()
# here, g = 1 creates a single random-effect group for the covariance
    structure
```

```r
data$g <- 1

VCV <- diag(data$vi, nrow = nrow(data))
rownames(VCV) <- colnames(VCV) <- data$Effect_id

# ensure the effect-size identifier is treated as a factor
data$Effect_id <- as.factor(data$Effect_id)

fit_tmb1 <- glmmTMB(yi ~ 1 +
                      equalto(0 + Effect_id|g, VCV) +
                       (1|Study_id) +
                       (1|Species_nonphylo) +
                       propto(0 + Species_phylo|g, A),
                       data = data,
                       REML = T)

# alternative way:
fit_tmb2 <- glmmTMB(yi ~ 1 + equalto(0 + Effect_id|g, VCV) +
                          (1|Study_id) +
                          (1|Effect_id) +
                          (1|Species_nonphylo) +
                          propto(0 + Species_phylo|g, A),
                          dispformula= ~ 0,
                          data = data,
                          REML = T)
```

In `brms`, sampling variances are incorporated via `se(sqrt(vi))`, which tells the model that each observed effect size has a known sampling standard error $\sqrt{v_i}$. This ensures that each effect size is modelled in the standard meta-analytical way, $y_i \sim \mathcal{N}(\mu_i, v_i)$ (eq.3). The phylogenetic effect is included as a group-level term with `gr(..., cov = A)`, where `A` is the phylogenetic covariance matrix.

You can find more detailed settings and also other alternative ways to use `brms` to fit a meta-analysis in our online tutorial.

```r
formula_phylo_brms <- bf(yi|se(sqrt(vi)) ~ 1 +
                          (1|Study_id) +
                          (1|Effect_id) +
                          (1|Species_nonphylo) +
                          (1|gr(Species_phylo, cov = A)))

prior <- get_prior(formula = formula_phylo_brms,
                   data = data,
                   family = gaussian())

fit_brms <- brm(formula = formula_phylo_brms,
                data = data,
                data2  = list(A = A),   # pass the phylogenetic correlation
                    matrix
                family = gaussian(),
                chains = 2,
                iter = 2000,
                warmup = 1000)
```

Having specified equivalent models in all three `R` packages, we compared the resulting estimates across packages. `metafor`, `glmmTMB`, and `brms` produced nearly identical overall effect estimates and

variance components when the same model was fitted. Because the `glmmTMB` results were effectively identical to those from `metafor`, Figure 3A shows only the `metafor` and `brms` results for visual clarity.

The estimated overall effect size was approximately 0.37 on the Fisher's Z scale, corresponding to about 0.35 after back-transformation to Pearson's r, indicating a moderate positive association consistent with assortative mating across taxa. Uncertainty intervals were also very similar across packages: 95% CI [0.11, 0.62] in both `metafor` and `glmmTMB`, and 95% CrI [0.04, 0.72] in `brms`.

Variance components were likewise highly consistent across packages. For within-study variation ($\sigma^2_{\text{Effect\_id}}$), all three packages gave the same point estimate of 0.01, with interval estimates of 95% CI [0.01, 0.02] in `metafor` and 95% CrI [0.01, 0.02] in `brms`. For among-study variation ($\sigma^2_{\text{Study\_id}}$), `metafor` and `glmmTMB` both gave a point estimate of 0.02, while `metafor` gave a 95% CI of [0.01, 0.03] and `brms` gave a very similar estimate of 0.02 with a 95% CrI of [0.01, 0.04]. For non-phylogenetic species-level variation ($\sigma^2_{\text{Species\_nonphylo}}$), point estimates were 0.06 in both `metafor` and `glmmTMB`, and 0.07 in `brms`; the corresponding intervals were 95% CI [0.03, 0.08] in `metafor` and 95% CrI [0.03, 0.08] in `brms`. For phylogenetic species-level variation ($\sigma^2_{\text{Species\_phylo}}$), all three packages gave the same point estimate of 0.05, with intervals of 95% CI [0.02, 0.18] in `metafor` and 95% CrI [0.02, 0.24] in `brms`. Thus, both phylogenetic and non-phylogenetic species effects contributed substantially to the total heterogeneity.

Total heterogeneity was very high, with identical estimates of $I^2_{\text{total}} \approx 97.3\%$ across all three packages, indicating that most of the variability among effect sizes reflected true heterogeneity across hierarchical levels rather than sampling error[15,23]. We calculated total I^2 as:

$$I^2_{\text{total}} = \frac{\sum \sigma^2_{\text{total}}}{\sum \sigma^2_{\text{total}} + \bar{v}} \times 100, \tag{11}$$

where $\sum \sigma^2_{\text{total}}$ is the sum of all heterogeneity components and $\bar{v}$ is the mean sampling-error variance. Based on the `metafor` fit, 38.6% of the variance was attributable to non-phylogenetic species effects, 35.5% to phylogenetic species effects, 13.3% to among-study effects, and 10.0% to within-study effects.

Phylogenetic heritability, defined as:

$$H^2_{\text{phylo}} = \frac{\sigma^2_{\text{phylo}}}{\sigma^2_{\text{phylo}} + \sigma^2_{\text{nonphylo}}}, \tag{12}$$

was estimated to be 0.479 (= 47.9%), indicating that about half of the among-species variance was explained by phylogenetic relatedness.

Key quantities to report include:

- the overall effect size together with its uncertainty (e.g. 95% CI or 95% CrI);
- the total heterogeneity (e.g. I^2) and its partitioning across model components;
- for phylogenetic models, the proportion of among-species variance explained by phylogeny (i.e. phylogenetic heritability).

These results can be visualised using orchard plots for the overall effect and forest or posterior distribution plots for the variance components (Figure 3A).

Although BM is the default correlation structure in phylogenetic meta-analysis, it can be useful to consider alternatives such as the Ornstein-Uhlenbeck (OU) model. OU correlation matrices can be generated in `ape`[36] or `phytools`[37] for a user-specified value of α, the parameter controlling the strength of adaptation. However, these functions do not estimate α from the data.

Because the OU and spatial exponential correlation structures share the same functional form, α can be estimated within the `metafor` framework by fitting an exponential spatial meta-analytic model, as

illustrated in our online tutorial. That estimate can then be used to construct the OU correlation matrix and re-fit the phylogenetic meta-analysis under the alternative correlation structure. For example,

```
# A is BM-based correlation matrix, I is the identity matrix (matrix with
    diagonal of 1s)
I <- diag(1, nrow(A))
D <- I - A # distance-like matrix
rho <- m_spatial$rho # rho is estimated from the spatial exponential model
    (m_spatial)
alpha <- 1 / rho
A_OU <- exp(-alpha * D) # OU-based phylo correlation matrix

m_OU <- rma.mv(yi, vi,
            random = list(~1|Study_id,
                          ~1|Effect_id,
                          ~1|Species_phylo,
                          ~1|Species_nonphylo),
            R = list(Species_phylo = A_OU),
            data = data)
```

The OU model provided a better fit to the data than the BM model (AIC = 330.8 vs. 345.4) and yielded a more precise estimate of the mean effect size (0.35, 95% CI [0.28, 0.42] vs. 0.37, 95% CI [0.11, 0.62]). The two models also differed in how they partitioned species-level variance. Under the OU model, most of this variance was attributed to the phylogenetic component ($\sigma^2_{\text{Species}_{\text{phylo}}} = 0.10$), with the non-phylogenetic species component close to zero. Under the BM model, by contrast, variance was divided between phylogenetic ($\sigma^2_{\text{Species}_{\text{phylo}}} = 0.05$) and non-phylogenetic ($\sigma^2_{\text{Species}_{\text{nonphylo}}} = 0.06$) species-level components. This pattern is consistent with the contrasting assumptions of the two models: BM allows variance to accumulate along branches, whereas OU implies stronger decay in correlation with evolutionary distance.

4.2. Spatial meta-analysis

We illustrate the spatial meta-analytic model introduced in eq. 4 using an exponential kernel to account for between-site spatial autocorrelation in effect sizes. The example is based on the dataset compiled by Grau-Andrés et al.[38], which synthesised plant responses to intensified fire regimes globally. The database contains 2,361 effect sizes from 393 primary studies conducted across diverse climates, habitats, and fire regimes, and reports standardised responses in plant abundance, diversity, and fitness to increased fire frequency or severity. Each comparison is expressed as Hedges' d, quantifying vegetation responses to more intense fire regimes relative to historical or lower-severity conditions.

In this dataset, multiple effect sizes are reported per primary study, and each study is associated with a single sampling location. Observed effect sizes y_i are indexed by effect-size identifier i, with study membership denoted by $j[i]$. We decomposed variation in y_i into effect-level heterogeneity, spatially structured among-study heterogeneity, and known sampling error.

Geographic coordinates for each study location were projected to calculate straight-line (Euclidean) distances in kilometres, which were then used to construct the exponential spatial correlation matrix. The model can be written as:

$$y_i = \beta_0 + \text{Location}_{l_{j[i]}} + \text{Effect\_id}_i + \text{Sampling error}_i,$$

$$\text{Location}_j \sim \mathcal{N}(0, \sigma_l^2 \mathbf{M}),$$

$$\text{Effect\_id}_i \sim \mathcal{N}(0, \sigma_s^2),$$

$$\text{Sampling error}_i \sim \mathcal{N}(0, \nu_i),$$

where $\mathbf{M}$ denotes the spatial correlation matrix among sampling locations, constructed from the exponential kernel $C_{\exp}(d_{xy})$, where d_{xy} is the Euclidean distance between locations x and y and ρ is the range parameter controlling the rate of correlation decay. The parameter $\sigma^2_{\text{spatial}}$ (σ_l^2 in eq.4) captures the marginal variance of the spatially structured location-level random effect, reported as τ^2 in `metafor` and as $sdgp^2$ in `brms`. We estimated the meta-analytic model using `metafor`, specifying a spatially correlated location-level random effect with an exponential correlation structure. The spatial component was implemented in two equivalent ways: (1) by supplying a location-level distance (or correlation) matrix, or (2) by providing projected x–y coordinates directly. In both cases, we also included an additional effect-level random intercept (Effect_id_i) to capture residual heterogeneity among effect sizes. The corresponding R code is shown below.

```r
data$const <- 1 # add a constant term used as a dummy grouping factor

# project to a planar coordinate system and convert to kilometres
dat_sf <- st_as_sf(data, coords = c("longitude", "latitude"), crs = 4326)
dat_sf_proj <- st_transform(dat_sf, crs = 3857)

coords_m <- st_coordinates(dat_sf_proj)
data$x_km <- coords_m[,1] / 1000
data$y_km <- coords_m[,2] / 1000
coords_km <- cbind(data$x_km, data$y_km)
dist_matrix_euclid <- as.matrix(dist(coords_km))

# row and column names
rownames(dist_matrix_euclid) <- data$Effect_id
colnames(dist_matrix_euclid) <- data$Effect_id

# exponential kernel----
## (1) Use distance matrix
EXP_meta1 <- rma.mv(yi, vi,
                    random = list(
                      ~ 1|Effect_id,
                      ~ Effect_id|const),
                    struct = "SPEXP",
                    data = data,
                    dist = list(Location = dist_matrix_euclid))

## (2) Use the coordinates directly (Euclidean distance)
EXP_meta2 <- rma.mv(yi, vi,
                     random = list(
                      ~ 1|Effect_id,
                      ~ x_km + y_km|const),
                     struct = "SPEXP",
                     data = data)
```

We can fit the same model structure in `brms`, incorporating known sampling variances via `se(sqrt(vi))`, together with a random intercept for `Effect_id` and a spatial Gaussian process term, `gp(x_km, y_km)`.

```
formula_spatial_brms <- bf(yi|se(sqrt(vi)) ~ 1 +
                     (1|Effect_id) +
                     gp(x_km, y_km, cov = "exponential", scale = FALSE))

prior <- get_prior(formula = formula_spatial_brms,
              data = data,
              family = gaussian())

EXP_brms <- brm(formula = formula_spatial_brms,
              data = data,
              family = gaussian(),
              prior = prior,
              iter = 2000,
              warmup = 1000,
              chains = 2,
              control = list(adapt_delta = 0.95, max_treedepth = 15))
```

In `glmmTMB`, we specified the spatial exponential random effect as `exp(pos + 0|const)`, where `pos` encodes projected coordinates via `numFactor()`, and incorporated the known sampling variance-covariance structure via `equalto(0 + Effect_id|const, VCV)`.

```
data$const <- 1 # add a constant term used as a dummy grouping factor for
    the spatial model
data$Effect_id <- factor(data$Effect_id) # ensure Effect ID is treated as a
     factor (categorical)
VCV <- diag(data$vi, nrow = nrow(data))  # create a diagonal variance-
    covariance matrix with sampling variances
rownames(VCV) <- colnames(VCV) <- data$Effect_id # assign Effect IDs as row
    /column names for VCV

data$pos <- numFactor(data$x_km, data$y_km) # convert spatial coordinates
    into numeric factor form for spatial random effects

EXP_tmb1 <- glmmTMB(yi ~ 1 +
                     equalto(0 + Effect_id|const, VCV) + # model sampling
                         variance structure using the VCV matrix
                     exp(pos + 0|const), # spatial exponential correlation
                         structure based on coordinates
                     data = data,
                     REML = TRUE)
```

The direction and magnitude of the overall effect were consistent across modelling approaches: $\beta_{0,\texttt{metafor}} = \beta_{0,\texttt{glmmTMB}}$ = -0.33 (95% CI [-0.46, -0.21]), and $\beta_{0,\texttt{brms}}$ = -0.32 (95% CrI [-0.49, -0.14]). Estimates of effect-size-level variation were also similar across frameworks, with $\sigma^2_{\text{Effect\_id}}$ = 0.79 (95% CI [0.73, 0.87]) in `metafor`, $\sigma^2_{\text{Effect\_id}}$ = 0.79 in `glmmTMB`, and $\sigma^2_{\text{Effect\_id}}$ = 0.89 (95% CrI [0.85, 1.02]) in `brms`. Spatial variance components were likewise comparable: τ^2 (i.e. σ^2_l in eq. 4) was 1.23 (95% CI [1.01, 1.50]) in `metafor`, 1.23 in `glmmTMB`, and $\texttt{sdgp}^2$ = 1.30 (95% CrI [1.00, 1.66]) in `brms`. For visual clarity, Figure 3B shows only the `metafor` and `brms` results.

The estimated spatial range parameters differed across packages because each framework uses a different parametrisation of the exponential covariance structure. In `metafor`, the exponential kernel is written as $\mathrm{cor}(d) = \exp(-\rho d)$, where ρ is a decay parameter expressed in km^{-1}. In `brms`, the corresponding function is written as $\mathrm{cor}(d) = \exp(-d/\mathrm{lscale})$, where *lscale* is a length-scale parameter expressed in km. In `glmmTMB`, the exponential correlation is internally parameterised through an inverse range parameter. Under a common exponential form, ρ and $1/\mathrm{lscale}$ play analogous roles.

In our example, however, the estimated decay-scale parameters differed across packages: `brms` gave *lscale* = 127.84 km (95% CrI [85.01, 186.52]), whereas `metafor` and `glmmTMB` gave $\rho = 0.17$ km^{-1}. Confidence and credible intervals for these parameters were also wide. This pattern suggests that, although spatially structured heterogeneity was consistently detected, the exact decay scale was only weakly identified in this dataset.

Biologically, an exponential kernel implies that sites closer together tend to have more similar effect sizes, with similarity declining as geographic distance increases. Using the `metafor/glmmTMB` estimate as an illustration, the expected correlation would be $\exp(-0.17 \times 1) \approx 0.84$ for sites 1 km apart, $\exp(-0.17 \times 5) \approx 0.43$ for sites 5 km apart, and $\exp(-0.17 \times 10) \approx 0.18$ for sites 10 km apart. However, because the estimated decay parameter varied markedly across packages, these values should be interpreted as package-specific examples rather than as robust estimates of the underlying spatial decay scale.

Total heterogeneity was high ($I^2_{\mathrm{total}} = 83.6\%$), indicating that most variability among effect sizes was not attributable to sampling error. Partitioning I^2 across model components suggested that 32.7% of the total variance was attributable to unstructured effect-size-level heterogeneity, whereas 50.8% was attributable to the spatially structured random effect. Here, I^2 was computed using the mean sampling variance as the representative sampling-error variance.

For spatial meta-analysis, key quantities to report include:

- the overall effect size and its uncertainty;
- the total heterogeneity and its partitioning across model components;
- the estimated spatial variance component;
- the spatial range or decay parameter, together with a clear explanation of how that parameter is defined in the fitted software. Because software packages differ in how they parameterise spatial decay, authors should report not only the numeric estimate but also the spatial correlation function used and a clear interpretation of the corresponding parameter.

We should note that the purpose of both our worked examples is to illustrate how meta-analytic models behave when applied to empirical data, rather than to provide a biological interpretation of the magnitude of individual effect sizes or to evaluate the validity of the underlying dataset or its data-collection procedures. Notably, the dataset used in the spatial meta-analysis contains some extremely large standardised mean differences, with Hedges' d occasionally reaching values of 10 or −20. Such values are biologically implausible in most ecological and evolutionary contexts and most likely reflect artefacts arising from zero or near-zero within-group standard deviations, or from misreporting or misinterpretation of sampling variability in the original studies. We used this dataset as a worked example because it provides a publicly available, geographically explicit meta-analytic structure suitable for demonstrating spatial dependence across packages. In practice, however, publicly available meta-analytic datasets with both sufficient geographic information and an appropriate hierarchical structure for cross-package comparison were surprisingly scarce.

We also note that in spatial meta-analysis, model convergence may depend on the specific combination of dataset and package. In some cases, a model converged in one package but failed to converge in another, even when the same model was being fitted. When this occurs, results should be compared carefully across packages to assess which components are robust to implementation and which are not. In our experience, fixed-effect estimates and most variance components were generally consistent, whereas

estimates of the spatial random effect were more likely to be unstable or to encounter convergence problems. All corresponding meta-regression analysis of our worked example dataset, together with an example of cross-package differences in convergence, is provided in the online tutorial.

5. Future perspectives and practical recommendations

We have outlined a step-by-step approach to conducting phylogenetic and spatial meta-analyses, emphasising how non-independence among species and locations can be incorporated within a multilevel framework. Explicitly modelling these forms of dependence is not merely a technical refinement, but an important condition for obtaining reliable inference in meta-analysis, because choices about model structure, correlation assumptions, and variance decomposition directly affect both estimates and their uncertainty.

Although our illustrative examples focus on ecology and evolution, the framework we describe is broadly applicable across research fields. In ecology and evolution, phylogenetic non-independence is pervasive and increasingly recognised as essential to address, whereas spatial non-independence also arises but remains less commonly modelled in practice. Spatial dependence is similarly widespread in other fields, including epidemiology and the social sciences (**Table 1**), where meta-analyses often rely on spatially structured data without explicitly modelling this dependence. Wider adoption of spatially explicit meta-analytic approaches has strong potential to improve inference and reduce overconfidence.

At the same time, broader use of phylogenetic and spatial meta-analyses brings important practical challenges. These models depend on correlation assumptions and on the estimation of random-effect variance components, both of which may be only weakly supported when data are sparse or poorly structured. In particular, separating phylogenetic or spatially structured variation from other sources of heterogeneity requires sufficient replication at the relevant hierarchical levels. When levels are weakly replicated or strongly confounded, increasingly complex models may yield unstable variance estimates or a misleading impression of precision. In such cases, simpler model specifications and sensitivity analyses may provide more robust insight.

Modelling choices should be guided not only by statistical fit, but also by biological and spatial plausibility. Default assumptions, such as Brownian motion for phylogenetic correlations or exponential kernels for spatial correlations, provide useful starting points, but alternative structures may better reflect the underlying processes in some systems. As our examples illustrate, different correlation assumptions can materially affect variance partitioning and effect estimates, and should be evaluated with appropriate caution.

Careful data preparation and clear reporting are also crucial. Uncertainty in phylogenetic trees, branch lengths, or spatial coordinates directly affects correlation matrices and downstream inference. Explicit documentation of these choices, together with reproducible code and, where feasible, sensitivity analyses, is critical. However, the present paper focuses specifically on phylogenetic and spatial dependence and does not address in detail other important sources of non-independence in meta-analysis, including dependence among sampling errors, temporal correlation, and other data-related dependencies that may also influence estimates and uncertainty. Readers should keep this in mind, as these other forms of dependence are beyond the scope of this paper.

As public datasets increase, the responsibility for synthesising this information increasingly falls to meta-analysts. We encourage researchers to treat phylogenetic and spatial information, along with analysis code, as essential components of reproducible synthesis rather than optional supplements[39–41]. Doing so will strengthen transparency and reproducibility while supporting more reliable and informative meta-analyses across research disciplines.

6. Acknowledgements

We are grateful to Wolfgang Viechtbauer for his foundational contributions to meta-analytic methods and software development, which have greatly informed this tutorial. In particular, this work builds extensively on the conceptual and practical framework provided by the `metafor` and `metadat` packages.

7. Funding Statement

SN and AM were supported by the Canada Excellence Research Chair (CERC-2022-00074) and NSERC Discovery Grant (RGPIN-2025-04813). AMS is supported by an Australian Research Council Future Fellowship (FT230100240).

8. Competing Interests

CW is a contributor to the `glmmTMB R` package.

9. Data Availability Statement

All data and code can be found in the GitHub repository.

10. Author Contributions

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

The authors wrote the original manuscript and used OpenAI ChatGPT (GPT-5.4) solely to enhance the text's clarity and readability. All sections were developed by the authors, who retain full responsibility for the content.

References

1. Nakagawa S and Santos ES. Methodological issues and advances in biological meta-analysis. Evolutionary Ecology 2012;26:1253–74.
2. Gurevitch J, Koricheva J, Nakagawa S, and Stewart G. Meta-analysis and the science of research synthesis. Nature 2018;555:175–82.
3. Koricheva J, Gurevitch J, and Mengersen K. Handbook of meta-analysis in ecology and evolution. Princeton University Press, 2013.
4. Adams DC. Phylogenetic meta-analysis. Evolution 2008;62:567–72.
5. Lajeunesse MJ. Meta-analysis and the comparative phylogenetic method. The American Naturalist 2009;174:369–81.
6. Hadfield JD and Nakagawa S. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. Journal of evolutionary biology 2010;23:494–508.
7. Koricheva J and Gurevitch J. Uses and misuses of meta-analysis in plant ecology. Journal of Ecology 2014;102:828–44.
8. Chamberlain SA, Hovick SM, Dibble CJ, et al. Does phylogeny matter? Assessing the impact of phylogenetic information in ecological meta-analysis. Ecology Letters 2012;15:627–36.
9. Cinar O, Nakagawa S, and Viechtbauer W. Phylogenetic multilevel meta-analysis: a simulation study on the importance of modelling the phylogeny. Methods in Ecology and Evolution 2022;13:383–95.
10. Legendre P. Spatial autocorrelation: trouble or new paradigm? Ecology 1993;74:1659–73.
11. Maire A, Thierry E, Viechtbauer W, and Daufresne M. Poleward shift in large-river fish communities detected with a novel meta-analysis framework. Freshwater Biology 2019;64:1143–56.
12. Viechtbauer W. Conducting meta-analyses in R with the metafor package. Journal of statistical software 2010;36:1–48.
13. Brooks ME, Kristensen K, Van Benthem KJ, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal 2017.
14. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. Journal of statistical software 2017;80:1–28.
15. Senior AM, Grueber CE, Kamiya T, et al. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. Ecology 2016;97:3293–9.
16. Nakagawa S, Noble DW, Senior AM, and Lagisz M. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. BMC biology 2017;15:18.
17. Nakagawa S, Yang Y, Macartney EL, Spake R, and Lagisz M. Quantitative evidence synthesis: a practical guide on meta-analysis, meta-regression, and publication bias tests for environmental sciences. Environmental Evidence 2023;12:8.
18. Van den Noortgate W, López-López JA, Marín-Martínez F, and Sánchez-Meca J. Three-level meta-analysis of dependent effect sizes. Behavior research methods 2013;45:576–94.
19. Van Den Noortgate W and Onghena P. Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. Educational and psychological measurement 2003;63:765–90.

20. Williams C, Yang Y, Warton DI, and Nakagawa S. Modelling approaches for meta-analyses with dependent effect sizes in ecology and evolution: A simulation study. Methods in Ecology and Evolution 2025;16:2362–79.
21. Felsenstein J. Phylogenies and the comparative method. The American Naturalist 1985;125:1–15.
22. Cornwallis CK and Griffin AS. A guided tour of phylogenetic comparative methods for studying trait evolution. Annual Review of Ecology, Evolution, and Systematics 2024;55.
23. Yang Y, Noble DW, Spake R, Senior AM, Lagisz M, and Nakagawa S. A pluralistic framework for measuring, interpreting and decomposing heterogeneity in meta-analysis. Methods in Ecology and Evolution 2025;16:2710–25.
24. O'meara BC, Ané C, Sanderson MJ, and Wainwright PC. Testing for different rates of continuous trait evolution using likelihood. Evolution 2006;60:922–33.
25. Hansen TF. Stabilizing selection and the comparative analysis of adaptation. Evolution 1997;51:1341–51.
26. Butler MA and King AA. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The american naturalist 2004;164:683–95.
27. Jetz W, Thomas GH, Joy JB, Hartmann K, and Mooers AO. The global diversity of birds in space and time. Nature 2012;491:444–8.
28. McTavish EJ, Gerbracht JA, Holder MT, et al. A complete and dynamic tree of birds. Proceedings of the National Academy of Sciences 2025;122:e2409658122.
29. Upham NS, Esselstyn JA, and Jetz W. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS biology 2019;17:e3000494.
30. Jetz W and Pyron RA. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nature ecology & evolution 2018;2:850–8.
31. Nakagawa S, Lagisz M, O'Dea RE, et al. orchaRd 2.0: An R package for visualising meta-analyses with orchard plots. Methods in Ecology and Evolution 2023.
32. Rios Moura R, Oliveira Gonzaga M, Silva Pinto N, Vasconcellos-Neto J, and Requena GS. Assortative mating in space and time: patterns and biases. Ecology Letters 2021;24:1089–102.
33. Viechtbauer W, White T, Noble D, Senior A, and Hamilton WK. Metadat: meta-analysis datasets. R package version 1.4-0 2021:1–.
34. Williams C, McGillycuddy M, Drobniak SM, Bolker BM, Warton DI, and Nakagawa S. Fast phylogenetic generalised linear mixed-effects modelling using the glmmTMB R package. bioRxiv 2025:2025–12.
35. Williams C, McGillycuddy M, Brooks M, et al. Meta-analysis with the glmmTMB R package. 2026. arXiv: 2604.04084 [stat.CO]. URL: https://arxiv.org/abs/2604.04084.
36. Paradis E and Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019;35:526–8.
37. Revell LJ. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ 2024;12:e16505.
38. Grau-Andrés R, Moreira B, and Pausas JG. Global plant responses to intensified fire regimes. Global Ecology and Biogeography 2024;33:e13858.
39. Parker TH, Forstmeier W, Koricheva J, et al. Transparency in ecology and evolution: real problems, real solutions. Trends in Ecology & Evolution 2016;31:711–9.

40. Nosek BA, Alter G, Banks GC, et al. Promoting an open research culture. Science 2015;348:1422–5.

41. Moreau D and Wiebels K. Nine quick tips for open meta-analyses. PLoS Computational Biology 2024;20:e1012252.

42. Guillerme T. dispRity: a modular R package for measuring disparity. Methods in Ecology and Evolution 2018;9:1755–63.

43. Pennell MW, Eastman JM, Slater GJ, et al. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 2014;30:2216–8.

44. Solmi M, Radua J, Olivola M, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Molecular psychiatry 2022;27:281–95.

45. Robles P de, Fiest KM, Frolkis AD, et al. The worldwide incidence and prevalence of primary brain tumors: a systematic review and meta-analysis. Neuro-oncology 2015;17:776–83.

46. Singh P, Arora A, Strand TA, et al. Global prevalence of celiac disease: systematic review and meta-analysis. Clinical gastroenterology and hepatology 2018;16:823–36.

47. Allum N, Sturgis P, Tabourazi D, and Brunton-Smith I. Science knowledge and attitudes across cultures: A meta-analysis. Public understanding of science 2008;17:35–54.

48. Balaj M, Henson CA, Aronsson A, et al. Effects of education on adult mortality: a global systematic review and meta-analysis. The Lancet Public Health 2024;9:e155–e165.

49. Branigan AR, McCallum KJ, and Freese J. Variation in the heritability of educational attainment: An international meta-analysis. Social forces 2013;92:109–40.

50. Chu C, Buchman-Schmitt JM, Stanley IH, et al. The interpersonal theory of suicide: A systematic review and meta-analysis of a decade of cross-national research. Psychological bulletin 2017;143:1313.

51. Chua JH, Cheng CKT, Cheng LJ, Ang WHD, and Lau Y. Global prevalence of resilience in higher education students: A systematic review, meta-analysis and meta-regression. Current Psychology 2023;42:22645–63.

52. Stahl GK, Maznevski ML, Voigt A, and Jonsen K. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. Journal of international business studies 2010;41:690–709.

Box 1: Sample size, model complexity, and identifiability in phylogenetic and spatial meta-analysis

Hierarchical meta-analytic models require sufficient information at each random-effect level to estimate variance components separately. Identifiability problems arise when different levels of the hierarchy are supported by nearly the same observational units, such that the data provide little information to determine how variation should be partitioned among variance components[1,9].

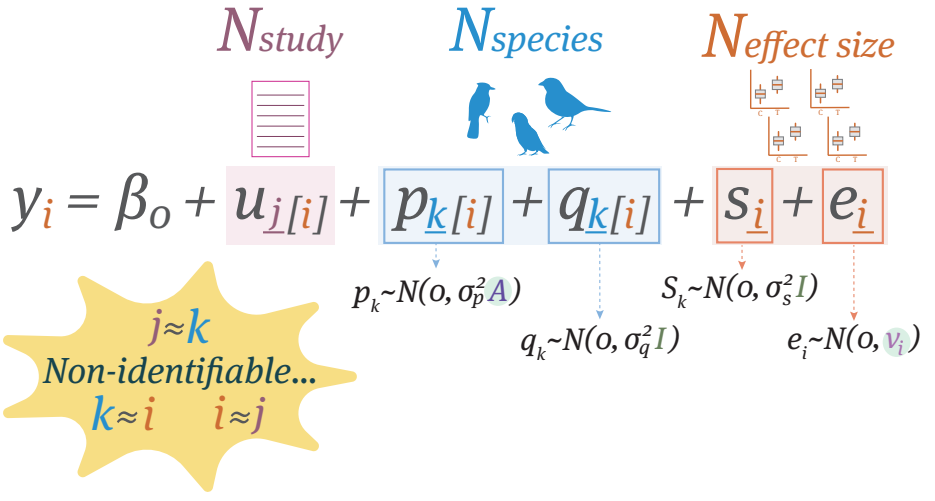

Figure B1. Schematic of a phylogenetic meta-analytic hierarchy illustrating how effect sizes, species, and studies are linked, and how near one-to-one mappings among these indices lead to weak identifiability of variance components.

Using Figure B1, we illustrate this issue using a phylogenetic meta-analysis model. Effect sizes (i) are the observational units, each associated with a study ($j[i]$) and a species ($k[i]$). Problems arise when these indices are nearly one-to-one mapped. For example, consider a meta-analytic dataset with 28 effect sizes from 21 species across 22 empirical studies, in which most species appear in only a single study and contribute only a single effect size. In this setting, study identity and species identity overlap strongly ($j \approx k$), providing little information to separate study-level from species-level sources of variation. Because most species also contribute only a single effect size ($k \approx i$), there is little information to distinguish species-level heterogeneity from effect-size–level heterogeneity. As a consequence, species-level variation is weakly informed by the data, making it difficult to further partition species-level heterogeneity into phylogenetically structured (p_k) versus non-phylogenetic (q_k) components.

More generally, each random effect contributes a variance parameter that multiplies a specified covariance structure. Sampling error is incorporated through known sampling variances, whereas unstructured random effects assume identity correlation structures, and phylogenetic or spatial random effects replace the identity matrix with correlation matrices derived from evolutionary or geographic distances. When the data do not contain enough information to distinguish among these structures, multiple random effects can explain similar patterns in the data, leading to weak

identifiability of individual variance components. Weak identifiability reflects limitations of the data structure rather than the absence of underlying biological processes.

In frequentist implementations, weak identifiability typically manifests as unstable variance estimates, with some components collapsing toward zero or variance being inconsistently allocated across components. In Bayesian implementations, weak identifiability is often reflected in poor chain mixing and strong posterior correlations among variance parameters, which can serve as diagnostic signals of limited information to support variance partitioning.

You might wonder why phylogenetic and non-phylogenetic species-level random effects (p_k and q_k) can, in principle, be distinguished despite operating at the same hierarchical level, and similarly why additional effect-size–level variation (s_i) can be separated from sampling error (e_i). In both cases, identifiability is enabled by known structure: phylogenetic effects are associated with a known correlation matrix (**A**), whereas sampling error variances (v_i) are fixed and known. When the data contain sufficient information, these known structures allow structured and unstructured sources of variation to be distinguished.

An analogous problem arises in spatial meta-analyses when sampling locations are nearly one-to-one with studies, and each location contributes only a single effect size. Under such conditions, spatially structured and study-level random effects are difficult to disentangle for the same reason. Identifiability in spatial meta-analysis depends not only on replication across locations but also on the information content of the spatial correlation matrix. When sampling locations are very close together, pairwise spatial correlations are uniformly high, providing little independent information to identify spatial structure. Conversely, when locations are very far apart, spatial correlations approach zero, leaving insufficient information to estimate spatial dependence reliably. In both cases, spatially structured random effects can explain patterns similar to those captured by other random components at the same hierarchical level, making it difficult to reliably partition variation among them.

Practical checks: Before fitting complex models, researchers should examine (1) the degree of replication at each hierarchical level (species/location and study) and (2) the distribution of effect sizes per species/location. When replication is sparse at hierarchical levels (e.g. species or location levels), or when hierarchical indices (e.g. species, location, and study) are nearly one-to-one mapped, simpler random-effect structures or sensitivity analyses may be more appropriate.

Box 2: Common issues in phylogenetic meta-analysis

The issues listed below arise when incorporating phylogeny into statistical models and are common across many applications, including but not limited to phylogenetic meta-analysis.

- **Missing branch lengths**
  Topological phylogenetic trees lack branch lengths. These can be approximated using Grafen's method (e.g. `compute.brlen()` in the package `ape`[36]), which scales nodes based on their relative depths.
- **Non-ultrametric (non–time-calibrated) tree**
  Sometimes, the phylogenetic tree you have is not ultrametric. This means the distances from the root to the tips are not the same for all species. We cannot use such trees directly to construct phylogenetic correlation matrices that assume a time-calibrated evolutionary history, because branch lengths reflect heterogeneous rates of evolution rather than elapsed evolutionary time. For phylogenetic models, we need a time-calibrated (ultrametric) tree. Ultrametricity can be enforced using algorithms such as `chronos()` in the package `ape`[36] or `force.ultrametric()` in the package `phytools`[37]. After converting a tree, check that the new branch lengths still make sense biologically. Also, record which method you used, since different approaches can change branch lengths and influence later estimates of phylogenetic heritability and variance components. In some cases, a tree that is intended to be time-calibrated may nevertheless be reported as non-ultrametric due to numerical precision issues or extremely small deviations in branch lengths. You can make small adjustments or use tolerance-based corrections to restore ultrametricity before analysis.
- **Polytomies**
  Unresolved nodes with more than two descendants can be converted to bifurcating form using tools like `multi2di()` in the package `ape`[36] (you can also do it manually - for example, based on literature). This process may introduce zero-length branches, which imply perfect correlation under Brownian motion. These branches can be replaced with small constants (e.g. 1e-8), fractions of typical non-zero branch lengths, or removed using functions such as `remove.zero.brlen()` from the `dispRity` package[42].
- **Taxonomic mismatches**
  Discrepancies between tree tip labels and species names in the dataset—due to synonyms, typographical errors, or outdated taxonomy—must be resolved before analysis. Functions like `name.check()` in the package `geiger`[43] can help identify such mismatches for manual correction.
- **Duplicated species**
  Some phylogenetic trees include multiple tips representing the same species (e.g. subspecies or hybrids). These duplications should be removed or merged, as correlation matrices typically assume one tip per species. If subspecies or hybrids are analysed as separate units, tip labels and dataset identifiers must be made explicitly consistent to avoid mismatches.
- **Missing species in the tree**
  Species present in the dataset may not be found in the tree. This may reflect unresolved taxonomic mismatches or true absence from the phylogeny. In the latter case, these species must be excluded unless their phylogenetic placement can be reliably inferred and added manually.
- **Extraneous species in the tree**
  Phylogenetic trees may include species that are not present in the dataset. In most implementations, tree tip labels must match the species identifiers used in the model exactly;

otherwise, model fitting will fail because the dimensions of the phylogenetic correlation matrix do not align with the data. Consequently, extraneous tips should be pruned manually or using functions such as `drop.tip()` or `keep.tip()` functions in the `ape` package[36] before analysis. Beyond preventing outright errors, pruning unnecessary tips is also important because their presence during earlier preprocessing steps (e.g. branch-length transformation or scaling under Grafen's method) can influence the resulting tree structure and downstream estimates of phylogenetic signal (e.g. λ).

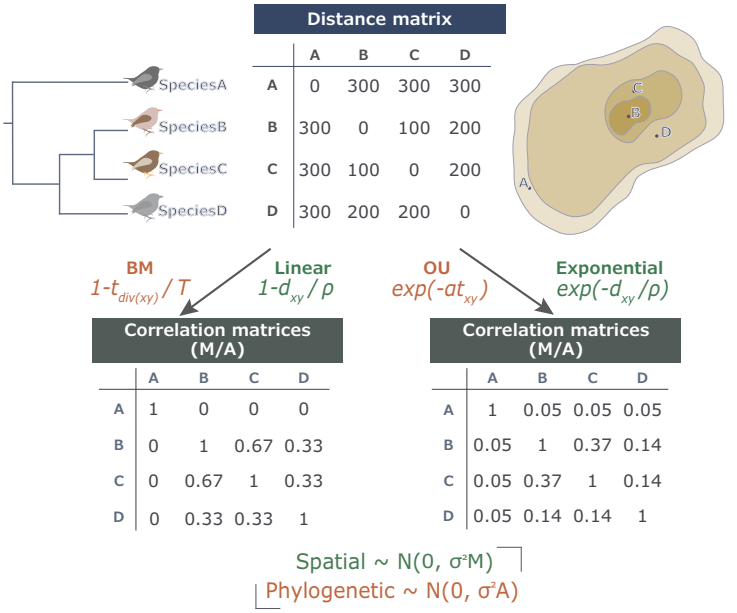

Figure 1: **Similarity between phylogenetic and spatial correlation structures in meta-analysis.** The figure illustrates how pairwise distances are transformed into correlation matrices and incorporated into random-effect structures in phylogenetic and spatial meta-analyses. Upper panels show example distance matrices derived from two alternative sources of non-independence. In the phylogenetic example (left), distances represent divergence times extracted from an ultrametric tree for four species (**A–D**), expressed here in thousands of years. In the spatial example (right), distances correspond to pairwise geographic distances among four sampling locations (**A–D**), expressed in kilometres. These two distance measures are not directly comparable in scale or interpretation: one represents evolutionary separation in time, whereas the other represents geographic separation in space. The letters indicate species or sampling locations, depending on the panel. In the spatial example, the shaded contours are included only as a schematic visual aid to convey relative proximity among locations. They do not represent real topography, spatial regions, sampling extents, or overlapping distributions. The example spatial distances are intentionally simplified for explanatory purposes and are not intended to depict a realistic geographic configuration. In particular, the contour-like shading is drawn such that locations **B**, **C**, and **D** are equidistant from **A**, mirroring the symmetric structure of the example spatial distance matrix and facilitating comparison with the phylogenetic case. The two lower correlation matrices are organised by correlation model rather than by data source. The left matrix shows the Brownian motion (BM) and linear-decay versions, whereas the right matrix shows the Ornstein–Uhlenbeck (OU) and exponential-decay versions. Within each matrix, the phylogenetic correlation matrix (**A**) and the spatial correlation matrix (**M**) are shown in parallel. For phylogenetic dependence, correlations are illustrated under (i) a Brownian motion (BM) process, where correlation between taxa x and y is given by $1 - t_{\text{div}}(xy)/T$, and (ii) an Ornstein–Uhlenbeck (OU) process with attraction parameter $\alpha = 0.001$, where correlation decays as $\exp(-\alpha t_{xy})$. For spatial dependence, correlations are illustrated under (i) a linear decay model, $1 - d_{xy}/\rho$, and (ii) an exponential decay model, $\exp(-d_{xy}/\rho)$, where d_{xy} denotes geographic distance and ρ is the range parameter. The linear spatial model is mathematically analogous to the BM model, whereas the spatial exponential model is equivalent to the OU model when $\alpha = 1/\rho$. Example correlation values are shown for units **A–D** under each model. Phylogenetic meta-analyses assume random effects distributed as $N(0, \sigma^2\mathbf{A})$, where **A** is the phylogenetic correlation matrix derived from the evolutionary tree. Spatial meta-analyses analogously specify $N(0, \sigma^2\mathbf{M})$, where **M** is the spatial correlation matrix derived from geographic distances. Together, the panels demonstrate that phylogenetic and spatial meta-analytic models rely on mathematically analogous transformations of pairwise distances into correlation structures, differing primarily in the definition of distance and the choice of kernel.

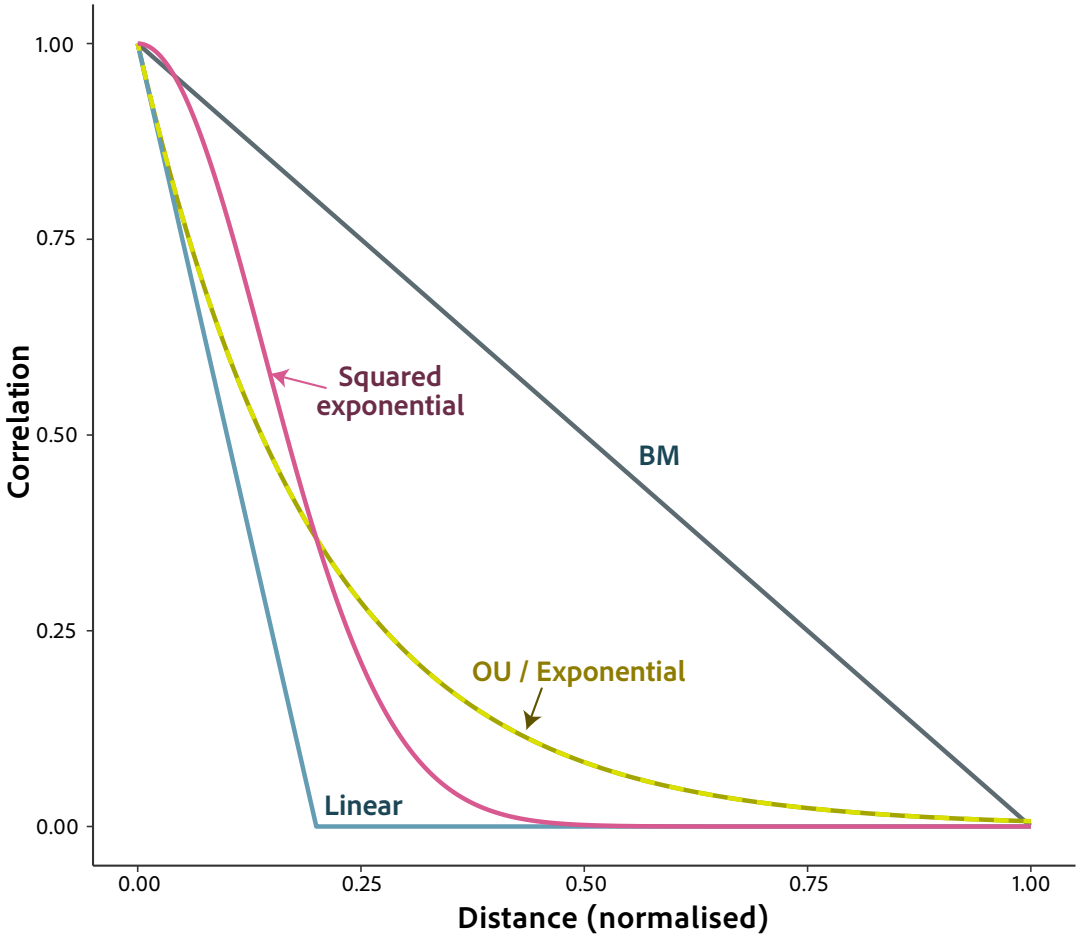

Figure 2: **Comparison of correlation-distance relationships for phylogenetic and spatial meta-analysis models.** Under a Brownian motion (BM) model, correlations decline linearly with evolutionary distance, which is analogous to a spatial decreasing linear kernel (function) model. The Ornstein-Uhlenbeck (OU) model, with $\alpha = 1/\rho$, is mathematically equivalent to the spatial exponential kernel. Spatial squared exponential (Gaussian) kernels are also shown to illustrate a commonly used spatial correlation structure that has no direct phylogenetic analogue. Unlike the linear (BM) and exponential (OU) models, which have clear counterparts across phylogenetic and spatial settings, the squared-exponential kernel yields an infinitely smooth, rapidly decaying correlation with distance and is typically restricted to spatial Gaussian process models.

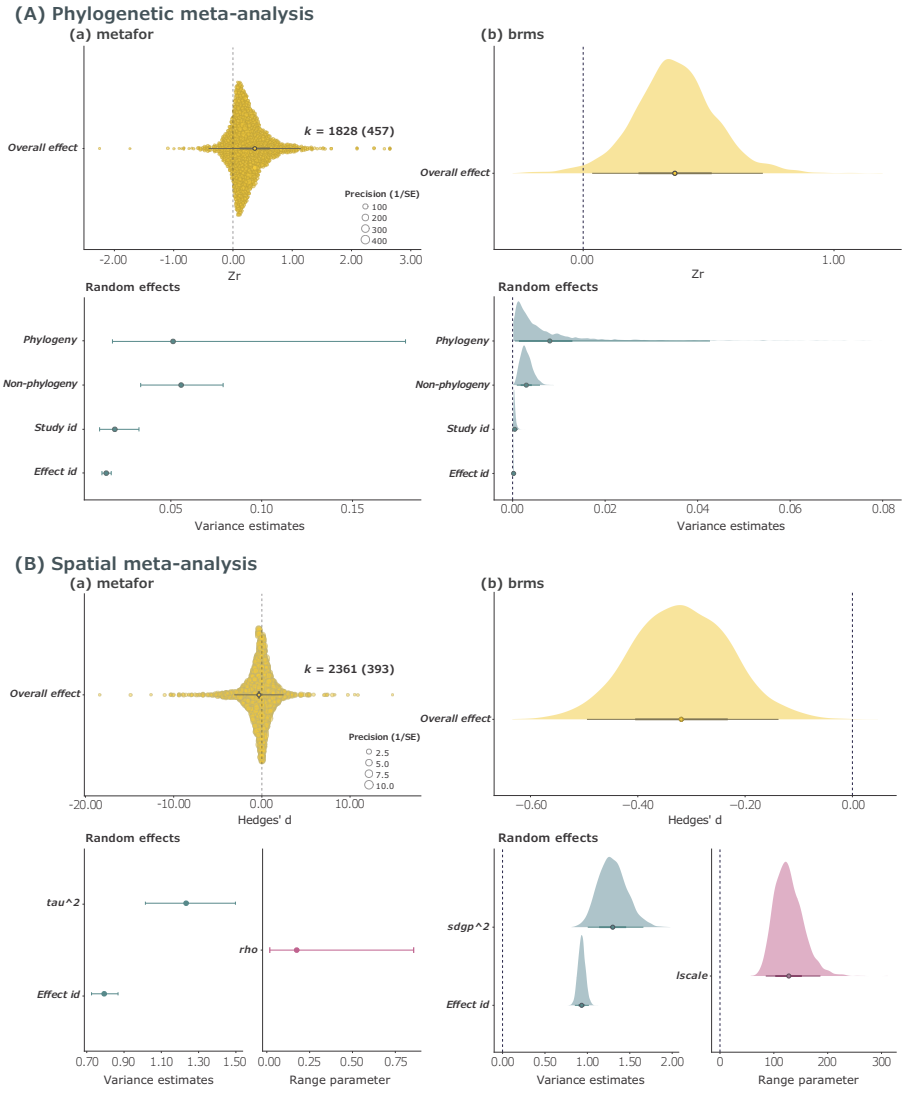

Figure 3: **Results of phylogenetic and spatial meta-analyses estimated using frequentist and Bayesian frameworks.** Panels (A) and (B) correspond to the phylogenetic and spatial meta-analyses, respectively; within each panel, subpanels (a) and (b) show results from frequentist (`metafor`) and Bayesian (`brms`) implementations of the same underlying model. (a) Frequentist results from `metafor`. Upper panels show orchard plots of the estimated overall effect size (expressed as Fisher's Zr for the phylogenetic analysis and Hedges' *d* for the spatial analysis). Individual circles represent effect sizes, with circle size scaled by precision (1/SE). The pooled estimate is shown as a filled circle in the centre, with thick horizontal lines indicating 95% confidence intervals and thin horizontal lines indicating 95% prediction intervals. Lower panels mainly display estimates of variance components, including random effects (phylogenetic and non-phylogenetic species effects for panel A; spatially structured study-level variance and also the associated range parameter for panel B). (b) Bayesian results from `brms` for the same models and datasets. Upper panels show posterior distributions of the overall effect size, with points indicating posterior medians and horizontal lines denoting the central 50% and 95% credible intervals. Lower panels show posterior distributions of the corresponding variance components and, for the spatial analysis, the spatial range parameter. The phylogenetic analysis is based on 1,828 effect sizes from 457 studies, and the spatial analysis on 2,361 effect sizes from 393 primary studies.

Table 1: Examples of meta-analyses outside ecology and evolution and how spatial meta-analysis could extend them. Spatial random effects capture geographic non-independence, thereby improving uncertainty estimates and quantifying the similarity of estimates within or between regions. However, identifying the drivers of such spatial patterns (e.g. cultural, genetic, or policy factors) requires additional explanatory variables

Discipline / Field	**Limitation (What is not done)**	**What spatial random effects could reveal**	**References**
Psychiatry	Estimates of age at onset are compared across countries, but without accounting for spatial autocorrelation between nearby regions.	More reliable uncertainty estimates and quantification of how similar onset ages are among nearby regions.	[44]
Neuro-Oncology	Incidence rates of brain tumours are reported globally, but neighbouring regions are treated as independent.	Identification of spatial dependence in incidence estimates and assessment of geographic ranges over which rates are similar.	[45]
Gastroenterology	The prevalence of celiac disease is compiled worldwide, yet regional non-independence is ignored.	Clarification of whether prevalence estimates show spatial dependence and how similarity decays with distance.	[46]
Sociology	Cross-national patterns in science knowledge–attitude links are compared, but regional similarity is not modelled.	Estimation of broader regional similarities (e.g. countries in Western Europe being more similar among themselves than to Eastern Europe).	[47]
Sociology	The effects of education on mortality or attainment are analysed country by country, without accounting for geographic correlation.	Quantification of whether countries form regions with stronger within-region similarity (e.g. high-income vs low-income blocs).	[48,49]
Psychology	Suicide theory constructs and resilience prevalence are compared across countries, but cross-national dependence is ignored.	Estimation of the degree of similarity within regions (e.g. Asian vs European student populations) relative to between regions.	[50,51]
Int. Business / Org. Psych.	Cultural diversity effects are meta-analysed across studies, but study location is treated as independent.	Assessment of how effects vary across geography and whether certain regions show stronger within-region similarity than others.	[52]