



Review and synthesis

Use of meta-analysis in forest biodiversity research: key challenges and considerations

Rebecca Spake^{a,b,*}, C. Patrick Doncaster^a^a Biological Sciences, University of Southampton, Southampton SO17 1BJ, UK^b Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK

ARTICLE INFO

Article history:

Received 7 March 2017

Received in revised form 15 May 2017

Accepted 22 May 2017

Keywords:

Biodiversity

Effect size

Forest

Meta-analysis

Review

Weighting

ABSTRACT

Meta-analysis functions to increase the precision of empirical estimates and to broaden the scope of inference, making it a powerful tool for informing forest management and conservation actions around the world. Despite substantial advances in adapting meta-analytical techniques for use in ecological sciences from their foundations in medical and social sciences, forest biodiversity research still presents particular challenges to its application. These relate to the long timescales of successional stages, often precluding experimental designs, and the often-large spatial scales required to select random plots for sampling treatment factors of interest. Empirical studies measuring biodiversity responses to forest treatments vary widely in their quality with respect to the number of treatment replicates and the randomness of their allocation to treatment levels, with a high prevalence of pseudoreplicated designs. It has been suggested that meta-analysis can potentially offer a solution to the vast pseudoreplicated literature, because results from pseudoreplicated studies are formative collectively. Here we review the principal issues that arise when including differently designed studies in meta-analyses of forest biodiversity responses to forest management or disturbance, in addition to more general matters of appropriate question formulation and interpretation of synthetic findings. These concern the need for questions of practical value to forest management, appropriate effect size estimation and weighting of primary studies that differ in study design and quality. We recommend against using effect sizes that are standardized against within-study variance when pooling studies across different designs or across factors such as taxonomic group. We find a need for alternative weighting schemes to the conventional inverse of study variance, to account for variation between studies in their design quality as well as their observed precision. Finally, we recommend caution in interpreting results, particularly with regard to the possibility of systematic biases between reference and treatment stands.

© 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction	430
2. Challenges and recommendations	431
2.1. Question formulation and the choice of an appropriate reference	431
2.2. Assessing study quality: selecting study designs to include in a meta-analysis	432
2.3. Choice of effect and estimation of its size	433
2.4. Weighting studies of varying quality	434
2.5. Interpretation of results	435
3. Conclusions	436
Acknowledgements	436
References	436

* Corresponding author at: Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK.

E-mail addresses: R.Spake@soton.ac.uk (R. Spake), cpd@soton.ac.uk (C.P. Doncaster).

1. Introduction

The primary response of conservation biologists to the rapid global loss of forest quality and extent has been to establish systems of protected areas that regulate against biodiversity loss. Whilst the strict protection of old-growth forests will likely remain a conservation priority throughout the world, the potential for other types of forests to support biodiversity is gaining increasing recognition (Gibson et al., 2011; Putz et al., 2008). The importance of diversity is recognized in the global-scale Strategic Plan for Biodiversity, drawn up by the Convention on Biological Diversity and agreed upon by the governments of the world in Aichi, Japan 2011. Target 11 of the Strategic Plan states that by 2020, at least 17% of the areas of particular importance for biodiversity and ecosystem services are to be conserved through “ecologically representative and well-connected systems of protected areas and other effective area-based conservation measures”. Target 7 advocates the implementation of sustainable management strategies for production forests, and Target 15 calls for the restoration at least 15% of degraded areas through conservation and restoration activities. If we are to affect these targets for the forest ecosystems of the world, we need a sound empirical and synthetic understanding of the functioning and the relative biodiversity value of forests under varying management regimes and conservation designations. A synthetic understanding is best achieved through the systematic collation of empirical research and meta-analysis of primary studies, which can yield practical generalizations for guiding forest management and conservation.

The number of published meta-analyses in forest biodiversity research has increased greatly over the last decade, keeping pace with empirical research in this field (Fig. 1). The aims of such meta-analyses vary widely, from seeking accurate estimates of a critical parameter of interest, such as a point estimate of the overall shape of published species-productivity curves (Whittaker, 2010), to attributing variation in effect size to meaningful covariates across a broad pool of studies (Lajeunesse, 2010). The former aim is generally explored with random-effects models, and the latter is achieved with multiple subgroup analyses or mixed-effects meta-regressions (Gurevitch and Hedges, 1999). Meta-analyses yield generalizations of practical value for informing forest management practice when they summarize the magnitude and direction of effect sizes that measure impacts on biodiversity and they

attribute variation in these effect sizes to meaningful covariates (Koricheva et al., 2013).

Meta-analysis was originally developed as a tool for the medical and social sciences, and was used extensively in these disciplines decades before its uptake in ecology (Lau et al., 2013). Systematic review and meta-analytical techniques have been adapted for use in ecology to account for higher empirical variability, necessitating different approaches to data synthesis (Pullin and Stewart, 2006), effect-size calculation (Lajeunesse, 2011), and meta-regression (Gurevitch et al., 2001). Despite substantial advances in the field, several authors (Halme et al., 2010; Koricheva and Gurevitch, 2014) have identified misuses of meta-analysis in ecology. Various guidelines exist to support ecological meta-analysis and interpretation (e.g. Harrison, 2011; Koricheva et al., 2013; Lortie et al., 2013), but the numerous recognized challenges have yet to be synthesised for applications to forest biodiversity. Meta-analyses of studies that measure biodiversity responses to forest management face particular issues to do with spatially configured plots measured over long timescales, and studies collated across diverse types and qualities of designs. Whilst these issues are not unique to forest biodiversity meta-analyses, they are particularly prevalent in this ecological discipline. For example, Gibson et al. (2011) meta-analyzed 2220 pairwise comparisons of biodiversity values in primary forests and disturbed forests that included studies with between 4 and 100 replicates and five different metrics of biodiversity. Chaudhary et al. (2016) meta-analyzed 1008 species richness differences between managed and unmanaged forests with between 2 and 336 replicates. Curran et al. (2014) meta-analyzed richness differences between old-growth and restored forest sites from 108 studies, with between 1 and 123 within-group replicates. The particular nature of the challenge is that replicates generally take the form of spatial plots, which make random and independent replication difficult at the forest scale. Moreover, the long timescales of successional stages often preclude experimental designs with precisely controlled treatment factors. The varying degrees of uncertainty among studies continues to impede our synthetic understanding of the ecological impacts of forest management and conservation interventions.

Here we focus on meta-analyses of biodiversity responses to forest management, disturbance, or conservation interventions. For each of the five principal stages of systematic review and meta-analysis, we identify key challenges for research on forest

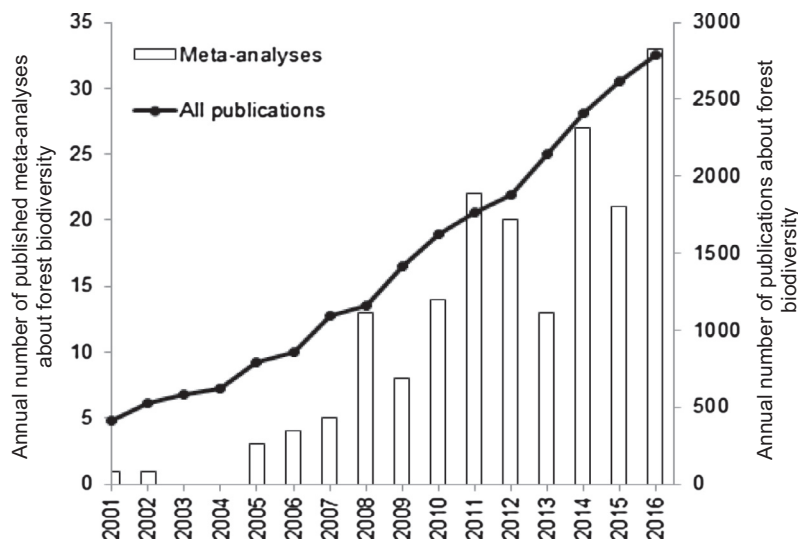


Fig. 1. Number of articles published per year in the ISI Web of Science containing the search terms ‘forest’ and ‘biodiversity’ (black lines) and also ‘meta-analysis’ (white bars).

Table 1

Key challenges and recommendations at the main stages of systematic review and meta-analysis for forest biodiversity (stages elaborated from Pullin and Stewart, 2006, and Nakagawa et al., 2017). See text for full explanations.

Stage of quantitative review	Key challenges for forest biodiversity	Recommendations
1. <i>Question formulation and choice of an appropriate reference</i> Determine a protocol for systematic review that defines (1) a test question with subject, intervention, and outcome elements, (2) the strategy for obtaining data, and (3) criteria for data inclusion or exclusion	Systematic reviews must formulate a test question of practical value to forest management Primary studies frequently have inconsistent definitions of old-growth or unmanaged forest, often allowing for inclusion of overmature plantations or forests heavily disturbed in the past	In landscapes with little remaining old-growth forest, consider alternatives to old-growth stands as the reference condition Prioritise consistency across studies in the characteristics of the reference stand, unless differences are meaningfully parameterised by moderators. State the limitations in knowledge about contributing studies
2. <i>Selection of study designs for inclusion in a meta-analysis</i> Critically appraise the extent to which research methodology is likely to prevent systematic errors or bias	Empirical studies vary widely in quality with respect to the number of treatment replicates and the randomness of their allocation to treatment levels, with a high prevalence of pseudoreplicated designs	Accommodate studies of varying quality and pseudoreplicated studies, which are informative collectively, with use of an appropriate effect-size estimator (stage 3) and weighting (stage 4)
3. <i>Choice of effect for estimation</i> Choose a common scale for meta-analyzing the results of multiple studies, by estimation of an 'effect size', representing the magnitude of a difference or the strength of a relationship	Effect sizes that are standardized against within-treatment variance (e.g., Hedges' <i>d</i>) may yield patterns that partially or only reflect differences in within-treatment variance, which depends on the study design (e.g., interspersions of treatment plots) and study scale (e.g., plot size and sampling extent) appropriate to the taxa and habitat types	Consider using a response ratio for the effect, as a more informative metric than a standardized effect in meta-analysis of studies with systematic differences in the magnitudes of within-treatment variance
4. <i>Weighting studies of varying quality</i> Weight study-level estimates of effect size according to study quality, conventionally by the inverse of effect-size variance to maximize precision of the meta-estimate	The lower effect-size variances generally obtained from pseudoreplicated designs, and that are sometimes obtained from little-replicated designs, give these studies a relatively large contribution to meta-estimates, despite their inferior quality	Consider alternative weightings to the inverse variance when including pseudoreplicated and little replicated ($n < 10$) designs. For example, ordinal classifications of study quality can account for study scale, treatment interspersions and replication
5. <i>Interpretation of results</i> Use meta-analyses to inform conservation policy and management only after cautious interpretation	The interpretation of effect sizes based on species density rather than species richness can be problematic in studies comparing biodiversity across forest types Two-level treatments often have confounding variation (e.g., unlogged stands located on steeper, higher, less productive slopes than logged stands)	Distinguish between species richness and species density, to account for potential influences of population abundances on effect size estimates Acknowledge all confounding variables associated with the selection of control and treatment stands that could have been consistent across empirical studies

biodiversity and propose recommendations for addressing them (summarized in Table 1).

2. Challenges and recommendations

2.1. Question formulation and the choice of an appropriate reference

Recommendation: Meta-analyses should make comparisons of practical value that are actionable to managers and conservationists. In regions with little remaining undegraded old-growth forest, consider using an alternative to old-growth or unmanaged forest stands as the reference condition. Whatever reference state is used, it must be used consistently across studies to guarantee the integrity of comparisons, unless differences can be parameterized by meaningful covariates.

Meta-analyses express the outcome of multiple studies on a common scale, through the calculation of an 'effect size', representing the magnitude of a difference or the strength of a relationship (Koricheva et al., 2013). Effect sizes that quantify the magnitude of a difference (such as mean differences or response ratios) require a reference against which to calibrate the treatment effect on the value of interest (such as biodiversity).

Empirical studies of forest biodiversity under a particular management regime typically measure the biodiversity value relative to its value in a reference state with attributes of an un-degraded ecosystem (Bullock et al., 2011). This is usually defined as 'old-growth', 'natural', 'undisturbed' or 'pristine' forest. The definition of old-growth varies considerably across regions, but such forests tend to be native to the region, with features of maturity, old age, primitive origins or lack of disturbance by modern technology (Beadle et al., 2009). Old-growth forest is widely used as a reference condition in forest biodiversity meta-analyses (e.g. Paillet et al., 2010; Chaudhary et al., 2016). A consistent application of this benchmark within a meta-analysis facilitates comparisons amongst empirical studies, with scope for synthesis of local responses to forest management at regional or global scales. Comparison with old-growth can moreover yield information on the capacity for managed stands or plantations to support equally rich communities including rare and dispersal-limited species. For example, meta-analysis of planted and secondary forests outside of tropical regions has demonstrated lower biodiversity than old-growth for functional groups with long recovery times from disturbance (Spake et al., 2015). Such studies lend support to calls for the exemption of old-growth from biodiversity offsetting schemes, and for initiatives to protect this declining and threatened habitat through reserve creation.

The choice of an undisturbed reference is not always appropriate, however, for analyses that seek generalizations of practical value; indeed, the concept of 'pristine' forest may no longer hold true in an era of pervasive anthropogenic change (Ghazoul et al., 2015). Although the strict protection of old-growth forests will likely remain a conservation priority throughout the world, the potential for other types of forests to support biodiversity is being increasingly realized (Gibson et al., 2011; Putz et al., 2008). Biodiversity and ecosystem service assessments must produce data relevant to management decisions affecting that management regime. It can be relevant to know the difference between the biodiversity value of forest under a management regime in its current state compared to a plausible alternative state, as opposed to an old-growth reference state. The alternative state might be a description of a possible future, based on the management question of interest (Peh et al., 2013). Meta-analyses that consider alternative reference conditions to old-growth or unmanaged forest stands can make relevant comparisons and produce results that are actionable to managers and conservationists in landscapes

where little old-growth forest remains, or where other forest management interventions or conversions are of interest. Given that meta-analysis is a resource-intensive undertaking, prioritisation of research questions is warranted, and we encourage future meta-analyses to make such comparisons. For example, Felton et al. (2010) compared the relative biodiversity value of plantation forests and pasturelands, as the majority of the world's plantations are generally afforested on former agricultural lands that are often of declining economic value for grazing or cropping.

In addition to selecting an appropriate reference stand for informing forest management and conservation, a key criterion for selection of a reference site is consistency across studies in the characteristics of the reference stand. A common criticism of meta-analysis is that researchers combine different kinds of studies ('apples and oranges') in the same analysis, leading to summary effects that ignore important differences across studies (Borenstein et al., 2009). Of particular relevance to forest-biodiversity research are inconsistencies in the definition of old-growth amongst the contributing studies for meta-analysis of a treatment compared to old-growth reference, often allowing for inclusion of overmature plantations or forests heavily disturbed in the past. Such reference stands can then encompass a diverse range of ages, land-use and disturbance histories. Although overmature plantations can support communities present in old-growth forest for some taxonomic groups (e.g. Spake et al., 2016), many studies have shown that this is not the case for other taxonomic groups, e.g. for understorey plants (Peterken, 1974). Meta-estimates of effect sizes that mix overmature plantations or heavily disturbed old forests with old-growth for comparison against younger secondary treatment forests will therefore fail to reflect true levels of biodiversity of old-growth forest. Using an inconsistent reference across studies is also likely to inflate effect-size heterogeneity between studies, due to differences between reference stands that are not parametrized by covariates in meta-analytic models. Because ecological studies always differ, the meta-analyst must seek a balance between not enforcing too-narrow a sampling universe (Lajeunesse, 2010) and still achieving meaningful effects from pooling across more-or-less similar studies (Stewart, 2010). Existing definitions of what constitutes old-growth forests can facilitate meaningful and consistent comparisons across studies (e.g. Frelich, 2002; Frelich and Reich, 2003; Wirth et al., 2009). Sound judgement is often impeded, however, by lack of information on the management histories of old-growth stands, and impaired by the potential value of expanding the domain of inference with inclusion of more studies (Ghazoul et al., 2015). We recommend that meta-analyses include clear statements on the limits of such knowledge about contributing studies (e.g. as in Curran et al., 2014).

2.2. Assessing study quality: selecting study designs to include in a meta-analysis

Recommendation: Pseudoreplicated studies, which have high prevalence in forest biodiversity research, provide collectively informative results, which can be accommodated by meta-analysis given appropriate decisions on the effect size estimator (Section 2.3) and the weighting on study quality (Section 2.4).

Empirical studies of forest biodiversity employ a range of study designs, with selection influenced by logistical, cost and time constraints, and the availability or spatial arrangement of forest stands across a landscape. Study designs vary in their capacity to minimize confounding variation. Some have argued that meta-analysis should be conducted on a select, homogeneous set of studies of the highest quality (e.g., Whittaker, 2010), whereas others favor a more inclusive approach that allows for broader understanding of ecological phenomena (Davies and Gray, 2015; Lortie et al., 2013). Such decisions can depend on the motive for doing the meta-analysis: the former increases power to detect effects and scope of inference; the latter is revealing about global effects that persist even across heterogeneous datasets (but not about those that do not).

Hurlbert (1984) distinguished 'acceptable' study designs, including completely randomized, randomized block and systematic designs all with independently interspersed treatment replicates, from pseudoreplicated designs, which lack independence across replicates (Fig. 2). More than 30 years after his seminal paper, pseudoreplication still features prominently amongst recent peer-reviewed forest biodiversity studies. Ramage et al. (2013) found definitive evidence of pseudoreplication in 52 (68%) of 77 studies published between 2001 and 2010 on logging in tropical forests; only 5 (7%) were definitively free of pseudoreplication, and the rest could not be clearly categorized due to poor reporting standards of study designs. Pseudoreplication remains widespread outside of the tropics also. Spake et al. (2015) found simple pseudoreplication in 15 (32%) of 47 studies comparing lichen, beetle and fungi species richness between old-growth and restored forest stands in temperate and boreal regions.

Pseudoreplication in forestry studies most commonly takes the form of simple segregation (Heffner et al., 1996). This involves analyzing multiple samples from a single contiguous treatment unit as if they were independent replicates interspersed amongst control replicates (Fig. 2). In empirical forest management studies, this can occur when treatment plots all come from one contiguous forest patch, and reference plots all come from another contiguous forest patch. Differences due to the treatment cannot then be unambiguously distinguished from other sources of spatial variation. The inability to disentangle these factors can result in

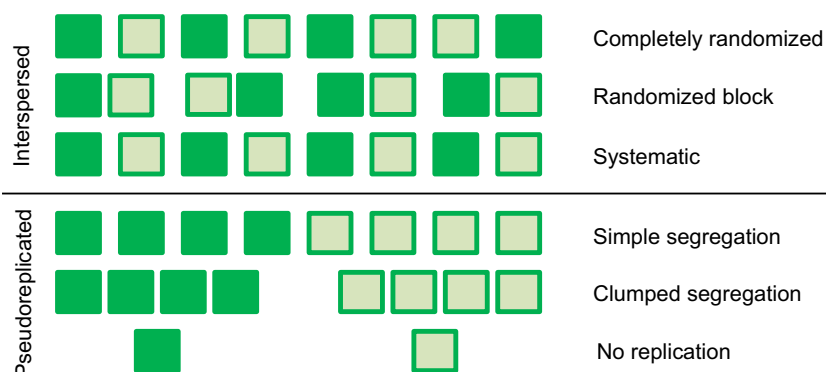


Fig. 2. Schematic representation of study designs for replicates (boxes) of two treatments (dark, light), often employed in empirical forest biodiversity research, and their classification into interspersed and pseudoreplicated sampling. Adapted from Hurlbert (1984).

spurious detection of apparent treatment effects or concealment of a real treatment effect (Ramage et al., 2013). When completely randomized designs are not possible, randomized block or systematic designs ensure that treatment units are interspersed rather than segregated (Fig. 2); within-block randomization reduces bias from other sources of spatial variation, and replication increases the precision of the estimate.

Several authors have considered the validity of including simply pseudoreplicated studies in meta-analyses. Halme et al. (2010) point out that such studies typically inflate standardized effect sizes, because of the low within-treatment variation, as well as confounding treatment effects with other sources of spatial variation. The ecological literature nevertheless includes repeated proposals for using meta-analysis to solve the problem of a pseudoreplicated literature (Hargrove and Pickering, 1992; Cottenie and De Meester, 2003; Davies and Gray, 2015). Proponents of this view argue that true patterns will emerge from the combined analysis of many independent studies, regardless of within-study pseudoreplication. This use applies to meta-analyses that focus on the direction rather than the magnitude of an effect, for example testing for a decrease in species richness across a management intervention (Ramage et al., 2013). Meta-analysis then becomes a statistical tool for randomly and independently replicating the studied systems to draw valid conclusions on the state of the population of systems (Cottenie and De Meester, 2003). The utility of pseudoreplicated studies in meta-analysis does not of course excuse drawing conclusions from a single pseudoreplicated experiment. It does however validate the publication of reports on large-scale experiments with inherent issues of replication due to the scale and complexity of the system (Cottenie and De Meester, 2003). Lajeunesse (2010) argued that sophisticated statistical approaches based on meta-regression techniques can facilitate the integration of studies of varying quality into a combined analysis. Regardless of developments in statistical sophistication, however, variation in design quality will always influence the choice of effect size and the weighting on primary studies. Here we argue that the principal problems concerning the integration of studies of varying quality relate to how the data are quantitatively synthesized, in terms of (i) what effect size is used and how it is calculated, and (ii) the use of weighting when pooling multiple study outcomes.

2.3. Choice of effect and estimation of its size

Recommendation: When including primary studies of varying design and quality, consider using a response ratio for the effect size, as a more informative metric than a standardized metric that conflates the magnitude of the difference in treatment means with the different magnitudes of random variation amongst studies.

One of the most widely used metrics in random-effects meta-analysis is the standardized mean difference. It is accurately estimated by Hedges' $d = J(\bar{x}_1 - \bar{x}_2) / S_{pooled}$, where \bar{x}_1 and \bar{x}_2 are the study means of the treatment and control groups respectively, S_{pooled} is the pooled standard deviation across the study, and J is a correction for low within-group replication (Hedges and Olkin, 1985). The standardization of a continuous response by S_{pooled} allows for a meaningful comparison to be made across studies with different residual variation (Rosenberg and Rothstein, 2013). Hedges' d is often chosen for ecological studies because it adjusts for differences amongst studies in sampling effort, corrects for small sample sizes, and can handle zero values for control or treatment groups (Rosenberg and Rothstein, 2013). The magnitude of Hedges' d can be difficult to interpret, however. Cohen's benchmark provides a rough classification of small, moderate, and large

effects from values of d in the region of 0.2, 0.5, and 0.8 respectively.

The calculation of S_{pooled} and J depends on whether the study has complete randomization, randomized blocks, or a systematic design, as does the estimate of variance in d (Borenstein et al., 2009; Kim and Becker, 2010). Meta-analyses in forest ecology rarely acknowledge this when describing effect-size calculations, leading to instances of the same calculation being used across completely randomized and simply segregated designs (as highlighted by Halme et al., 2010). Any form of pseudoreplication tends to inflate the magnitude of Hedges' d because the interdependence amongst replicates underestimates variability in the population.

Even if Hedges' d is applied to interspersed designs and is correctly calculated, its estimates may have dubious meaning for meta-analyses in forest ecology. The application of Hedges' d to a random-effects meta-analysis assumes that each primary study i has a value of $(S_{pooled})^2$ that estimates study variance σ_i^2 , which itself is drawn at random from the same population variation σ^2 across all studies. For many ecological meta-analyses (as distinct from the behavioral meta-analyses that originally inspired Hedges' d), differences between studies in spatial configuration, geographical distribution and taxonomic group will dictate against them having a common σ^2 (Osenberg et al., 1997). It is therefore possible, though almost universally overlooked, that cross-study differences in Hedges' d have less to do with treatment effects than differences in study design or taxonomic group. We therefore recommend against using standardized metrics for meta-analyses that collate studies of different design, scale or taxonomic group. The use of response ratios can overcome this problem (Gurevitch et al., 2001).

In addition to the spatial interspersion of treatment types, the study scale in terms of the study area extent and the size of sampling plots can affect estimates of pooled within-treatment standard deviation, and so Hedges' d (Fig. 3). For a completely randomized design, treatment and experimental units spread across a larger spatial extent will likely yield higher within-treatment variability. Indeed, whether studies are experimental or observational, those with sample plots dispersed across larger and more heterogeneous landscapes will be more exposed to co-varying factors such as soil type and climate confound the treatment effects, leading to greater variability among replicates (Osenberg et al., 1997; Monserrud, 2004). Smaller plot size leads to fewer individuals sampled per plot, and therefore sampled communities are probabilistically more likely to deviate from null and exhibit higher within-treatment variation (Chase and Knight, 2013). If within-treatment variation depends on study area extent and sample plot size, which can vary somewhat systematically across taxonomic group and habitat type, patterns in Hedges' d may emerge that have little to do with the actual strength of a process but, instead, reflect differences in within-treatment variation (Osenberg et al., 1997). The log response ratio provides an alternative and increasingly popular metric in ecological meta-analysis, which does not require within-group variance for its calculation. It measures $\ln(\mu_1/\mu_2)$; for each study i , estimated by $\ln R = \ln(\bar{x}_1/\bar{x}_2)$. The meta-estimate of $\ln R$ across a population of studies can be back-transformed to a percentage differences between control and treatment groups, as an intuitively accessible effect size for policy-makers. The calculation of variance in $\ln R$ depends on the study design. A completely randomized design has: $Var_{\ln R} = s_1^2/(n_1\bar{x}_1^2) + s_2^2/(n_2\bar{x}_2^2)$. A matched-pair (i.e., randomized-block) design requires a further subtraction from this estimate, of an amount equal to $2rs_1s_2/(\bar{x}_1\bar{x}_2n)$, where r is the coefficient of correlation between treatment and control observations and n is the number of matched pairs (Lajeunesse, 2011). For studies of forest biodiversity in particular, randomized blocks or systematic groups

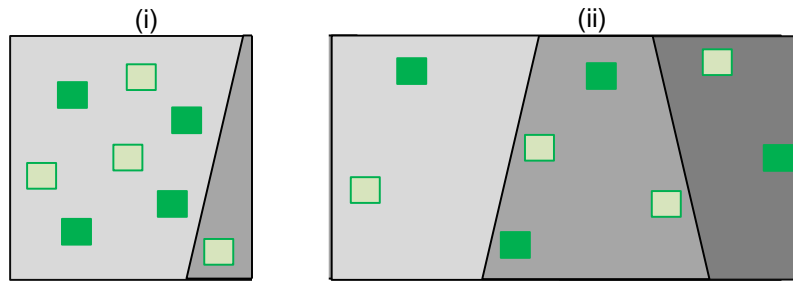


Fig. 3. Schematic representation of completely randomized designs testing the biodiversity difference between two treatments (dark, light squares) that cover two different extents (i and ii) across a landscape that is heterogeneous in soil type (indicated by different shades of gray). The design with plots dispersed over a larger extent (ii) will likely yield higher within-treatment variability due to variation in soil types across replicates.

of forest stands generally have low replication, which makes for nonsense estimations of r , and consequently of variances. Clearly further developments in effect-size metrics are needed for little-replicated studies, such as those prevalent in forest biodiversity.

An often-used motive for favoring Hedges' d over $\ln R$ is that $\ln R$ cannot be calculated when a treatment or control mean is equal to zero, which can occur when comparing biodiversity values across treatments of less speciose taxa. A common solution to this problem is to add small values, either integers or percentages to the zero values (e.g. Kalies et al., 2010). This is the default setting in the widely used R statistical package 'metafor' (Viechtbauer, 2010). Such additions can result in abnormal estimates of effect size. This problem is mitigated by sensitivity analysis of effect sizes, comparing outcomes of multiple trials of integer or fractional additions to zero values and ascertaining which addition least impacts on average effect sizes (Kalies et al., 2010; Molloy et al., 2008). Moreover, Lajeunesse (2015) derived estimators that correct for biases that arise when quantifying response ratios for studies with small sample sizes.

2.4. Weighting studies of varying quality

Recommendation: When primary studies vary markedly in their design, as they typically do in meta-analyses of forest biodiversity, consider alternative weightings to the conventional inverse of the effect size variance. Pseudoreplicated studies typically have low variance, and designs with very low replication may do as well, which can be more fairly accommodated in weightings by ordinal classifications.

Meta-analysis conventionally weights the estimate of effect size from each primary study by the inverse of its variance, in order to maximize the precision of the meta-estimate of effect size (Hedges and Olkin, 1985; Koricheva and Gurevitch, 2014). This inverse-variance weighting is considered a hallmark of meta-analysis, and it features on a checklist of quality criteria for ecological meta-analysis by Koricheva and Gurevitch (2014). The inverse-variance weighting is generally recommended over unweighted models when sufficient data are reported to compute standard errors of individual effect size estimates, and when parametric assumptions are not seriously violated (see Gurevitch and Hedges, 1999).

Many ecological meta-analyses nevertheless report unweighted meta-estimates, including approximately one quarter of those in plant ecology (Koricheva and Gurevitch, 2014). Authors justify the use of unweighted meta-analyses largely on the principle that it would unnecessarily reduce the number of primary studies to include only those with sufficient data for estimating variance (Koricheva and Gurevitch, 2014; Stewart, 2010). Gurevitch and Hedges (1999) concluded that some information regarding the overall findings is much better than no information. They called

for the development of methods for data synthesis of poorly reported data (e.g., where no estimate of sampling variance is published). Various solutions have been suggested, including weighting by an ordinal classification of study quality (e.g. Vanderwell et al., 2007), nonparametric weighting (Adams et al., 1997; Mayerhofer et al., 2013), imputation methods that fill in missing data (Lajeunesse, 2013) and Bayesian models (Ogle et al., 2013). A general rise in the standards of reporting primary studies seems likely to render this issue less problematic in the future (Koricheva and Gurevitch, 2014; Gerstner et al., 2017).

Meta-analyses that use the conventional weighting must address further issues that arise if primary studies include different taxonomic groups or different designs. Comparisons across taxonomic groups may require separate subgroup analyses. Mixed designs can be accommodated by correctly structuring the statistical model. Alternatively, phylogenetically controlled meta-analysis can not only deal with issues of non-independence (effect sizes derived from closely related species may be similar in magnitude or direction due to shared evolutionary history), but also the fact that different variances are expected because lineages within phylogenies may have evolved at different rates (Lajeunesse et al., 2013). Pseudoreplicated designs present special problems, however. The lower effect-size variance generally obtained from pseudoreplication will afford these studies a relatively large contribution to the meta-estimate, despite their inferior design. None of the pseudoreplicated designs in Fig. 2 have any true replication with which to measure variation. It may then make more sense not to weight, or to weight by the true sample size of independent replicates (Mayerhofer et al., 2013) on the principle that precision increases with sample size, all else being equal. Pullin and Stewart (2010) suggests using sensitivity analyses to investigate the impact of pseudoreplication on meta-estimates. Similar considerations arise with the inclusion of little-replicated studies. Designs with less than 10 replicates become susceptible to large errors in the estimation of variance (Hedges and Pigott, 2001), which can be investigated with sensitivity analysis and accommodated with a weighting on replication.

Estimation of the between-study variance, represented by the variance of the distribution of the true study effects (denoted as τ^2 in the meta-analytic literature), is often a core output of a meta-analysis, but also comprises an intermediate step in fitting a conventional random-effects model (Veroniki et al., 2016). In general, τ^2 will increase as the variance within studies decreases and/or the total observed variance increases. The inverse variance weighted random-effects meta-analysis estimates a summary treatment effect as the weighted average of study-wise effect sizes, with weights depending on both within-study (sampling) variance and the estimated between-study variance. Consequently, the estimated amount of the between-study variance influences the weights assigned to each study and hence the overall summary

treatment effect and, importantly, its precision (Veroniki et al., 2016). If the estimate greatly exceeds the within study random variation, the meta-analysis will have little potential to fulfil its principal function of improving the precision of an effect-size estimate.

Meta-analysis needs better methods and weighting schemes for synthesizing studies that vary markedly in their design and quality. As multiple study factors can work simultaneously to increase and decrease effect size variance, we consider that the development of ordinal weighting schemes, utilizing study quality criteria relevant to forest biodiversity studies, holds the most promise for addressing weighting issues. No off-the-shelf checklists exist for study quality that will be applicable to all forest biodiversity meta-analyses, and alternative designs generally do not fall into a clear quality hierarchy (Pullin and Stewart, 2006). Meta-analysts may wish to consult previous quality assessment tools to identify factors to code for study quality (reviewed by Bilotta et al., 2014). Here we have described how various factors affect study quality including the interspersed treatment stands across space, study area extent and the size of sampling unit and replication, which could be used to code for quality.

2.5. Interpretation of results

Recommendation: We urge caution in the interpretation of results, particularly with regard to recognizing the possibility of systematic biases between reference and treatment stands across studies, and clearly distinguishing between species richness and species density.

The great advantage of meta-analyzing a systematic review lies in the capacity to quantify a meta-effect of interest and to control bias in its interpretation. Yet subjective judgement is required at many stages of the process. Due care must therefore be taken with respect to the interpretation of results, because quantitative syntheses have potential to inform conservation policy and forest management.

Species richness is the simplest and most widely used measure of forest biodiversity (Magurran, 2004). For purposes of comparing species richness among sites or treatments, researchers may (i) adjust or control for differences in the number of individuals and the number of samples collected (rarefaction), or (ii) use abundance or incidence distributions to model the number of undetected species (estimators of asymptotic richness) (Gotelli and Colwell, 2001). However, most empirical studies of spatial variation in species richness standardize the richness measure against area or sampling effort, by employing an equal sampling intensity across sites. Such comparisons are actually comparisons of species density: the number of species per unit area (Gotelli and Colwell, 2001), or 'within-sample species richness' (Newbold et al., 2015). The use of species density rather than species richness (rarefied or estimated) can be problematic in studies comparing biodiversity across different forest types. It is possible that differences in species richness between, for example, old-growth forest and secondary forest stands could be partly driven by effects on overall abundance of individuals. The distinction between species richness and density is often not clarified in published meta-analyses, ignoring the influence of population abundances on richness, and we call for studies to make this distinction in the future (e.g., as in Newbold et al., 2015).

Whether species density or species richness is used to estimate the magnitude of a difference between different forest treatments, the effect size depends critically on two factors: (i) the size of the species pool, and (ii) sampling grain/extent (Chase and Knight, 2013). The size of a species pool might influence the measured effect sizes regardless of the true magnitude of the effects (Chase and Knight, 2013). A standardized sampling protocol, such as

area-controlled (species density) or individual- or sample-controlled (rarefaction) will sample a much smaller proportion of a community with a large species pool compared to a community with a smaller species pool (Chao and Jost, 2012). This can lead to underestimating the true biodiversity difference between e.g. old-growth and secondary forests. A higher intrinsic species richness in old-growth forests leads to a systematic undersampling bias that misses more species per site than in secondary-growth sites. In a global meta-analysis of old-growth and secondary forest biodiversity differences across a range of taxa, Curran et al. (2014) assessed the magnitude of undersampling bias by comparing sample completeness, in terms of the ratio of observed and estimated species-richness values among old-growth and secondary forests. They found that undersampling was prevalent across all studies and forest types, and concluded that the highly variable trend was unlikely to exert a serious bias toward one forest type relative to the other. They found that imbalances in sampling effort across forest types did influence effect-size calculation, and emphasized the need for more intensive surveys and replication (Curran et al., 2014).

Study scale, in terms of sampling plot size and study area extent, can influence the magnitude of effect-size estimates through its impact on effect-size variance (as described above) and directly on the effect-size estimate. For example, old-growth forest is expected to exhibit higher levels of spatial species turnover or beta diversity relative to secondary forest at different stages of growth; turnover is often more strongly influenced by geography (e.g., differences in topography or soil) in old-growth than secondary forest, where local conditions and disturbance history effects can dominate (Gardner et al., 2007). It is therefore possible that a higher intrinsic rate of spatial turnover of species in old-growth forest could lead to biased biodiversity comparisons with secondary forests when site replication or spatial scale of sampling is limited (Curran et al., 2014).

With species pool size (undersampling bias) and study area extent potentially biasing effect-size estimates of biodiversity differences between forest types, much caution is warranted in interpreting effect-size covariates. Many meta-analyses find systematic variation in effect sizes among covariates such as taxa (e.g. animals vs. plants) or biogeographic regions (e.g. temperate vs. tropical). This could in fact be due to differences in the size of the species pool or systematic variation in sampling scale among these comparisons rather than any inherent ecological differences (Chase and Knight, 2013). It is also important to note that using species richness alone as a proxy for biodiversity can be misleading when treating it as a decision variable for setting conservation and management priorities, as the metric takes no account of the composition of communities with regards to rare species or functional traits. With local and regional assemblages experiencing a substitution of their taxa, rather than systematic loss (Dornelas et al., 2014), there is a need to expand the focus of research and planning from biodiversity loss to community change.

The majority of forest biodiversity comparisons in empirical and meta-analytic research compare richness at the scale of forest stands, and are therefore comparisons of alpha diversity. Changes in human land-use are increasingly modifying the landscape-scale distribution of forest stands of different ages, dominant canopy species and management intensity. Forest management for biodiversity should not prioritize sites only on the rankings of their alpha diversity; regional-scale impacts on biodiversity within the context of heterogeneous landscapes also requires consideration (Chaudhary et al., 2016). Quantification of the degree of compositional differentiation among stands, as measured by beta diversity, enables a better understanding of the relative importance of stand-level forest management, limited dispersal and local stochastic processes.

Empirical studies and any meta-analyses that synthesize data from them must acknowledge all confounding variables associated with the selection of control and treatment stands. Confounding can arise even in designs with fully interspersed treatment and control stand. For example, in studies of logging impacts on biodiversity, unlogged control forest stands are usually located in steep, high elevation sites with generally less fertile soils and lower productivity than logged stands. Such differences are likely to have consequences for stand structure and biotic communities independently of any logging treatment applied to such stands, reducing the likelihood of detecting significant harvest impacts on biodiversity (Lindenmayer and Laurance, 2012). Further difficulties arise with quantifying cumulative impacts of logging across space or over time (Lindenmayer et al., 2012; Lindenmayer and Laurance, 2012). Low intensity selective logging can have little impact after a single cycle, while repeated logging events can exacerbate biodiversity impacts (Gibson et al., 2011; Woodcock et al., 2011; Edwards et al., 2011; Lindenmayer and Laurance, 2012). Past management intensity can be confounded with stand age, often with younger stands experiencing a higher number of rotations, which then requires acknowledgement in any meta-analysis of relationships between stand age and biodiversity. Old-growth forests often include legacies of past disturbances about which we have little or no knowledge (Ghazoul et al., 2015). It is possible that old-growth stands themselves may comprise a depauperate subset of the biotic community that was once supported in the past. A meta-analysis that fails to acknowledge past disturbance would then under-estimate the added value of old-growth over currently disturbed forest. The selection of contributing studies to the meta-analysis therefore merits careful consideration of the past disturbance context and potentially confounding differences amongst control and treatment stands. Gardner et al. (2007) outline sources of bias to consider that can lead to underestimates of biodiversity differences between tropical primary and secondary forests. Felton et al. (2010) identified four generalizations from meta-analysis results that depend on multiple factors: (i) landscape context, (ii) management practices, (iii) the conservation value of the taxa being considered, and (iv) the evaluated components of biodiversity. Great caution is warranted in making general statements about the relative biodiversity benefits of one forest type over another, particularly when the meta-analysis has not included confounding variables as covariates or factors in the model.

3. Conclusions

Meta-analysis of empirical forest biodiversity can yield practical generalizations for guiding forest management and conservation. Empirical studies in this field vary widely in their quality with respect to the number of treatment replicates and their spatial interspersion, with an ongoing prevalence of pseudoreplicated designs. Meta-analysis can accommodate a pseudoreplicated literature, if it focuses on the direction rather than the magnitude of effects. Numerous problems remain, however, with integrating studies of varying quality into a meta-analysis. Standardized measures of effect size are particularly prone to bias when pseudoreplication is an issue, and the log-response ratio for matched-pair designs has no sensible estimate of effect-size variance for studies with little replication. These issues have relevance to the application of the conventional inverse-variance weighting on primary studies in a meta-analysis. Finally, we highlight the need for cautious interpretation of generalizations from published and future meta-analyses, particularly with respect to the biodiversity metric used and the potential for confounding variation. These findings can serve as a guide for future meta-analyses in forest biodiversity research.

Acknowledgements

We thank A. Purvis for informative discussion of concepts and G. Petrokofsky for comments on an earlier version of the manuscript. This work was supported by grants to RS from the Biotechnology and Biological Sciences Research Council (grant number BB/H531935/1) and the Japanese Society for the Promotion of Science.

References

- Adams, D.C., Gurevitch, J., Rosenberg, M.S., 1997. Resampling tests for meta-analysis of ecological data. *Ecology* 78, 1277–1283.
- Beadle, C., Duff, G., Richardson, A., 2009. Old forests, new management: the conservation and use of old-growth forests in the 21st century preface. *For. Ecol. Manage.* 258, 339–340.
- Bilotta, G.S., Milner, A.C., Boyd, I.L., 2014. Quality assessment tools for evidence from environmental science. *Environ. Evidence*, 3, 14.
- Borenstein, M.B., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. *Introduction to Meta-analysis*. John Wiley, Sons, Ltd., Chichester, UK.
- Bullock, J.M., Aronson, J., Newton, A.C., Pywell, R.F., Rey-Benayas, J.M., 2011. Restoration of ecosystem services and biodiversity: conflicts and opportunities. *Trends Ecol. Evol.* 26, 541–549.
- Chao, A., Jost, L., 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93, 2533–2547.
- Chase, J.M., Knight, T.M., 2013. Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecol. Lett.* 16, 17–26.
- Chaudhary, A., Burivalova, Z., Koh, L.P., Hellweg, S., 2016. Impact of forest management on species richness: global meta-analysis and economic trade-offs. *Nature* 6, 23954.
- Cottenie, K., De Meester, L., 2003. Comment to Oksanen (2001): reconciling Oksanen (2001) and Hurlbert (1984). *Oikos* 100, 394–396.
- Curran, M., Hellweg, S., Beck, J., 2014. Is there any empirical support for biodiversity offset policy? *Ecol. Appl.* 24, 617–632.
- Davies, M.G., Gray, A., 2015. Don't let spurious accusations of pseudoreplication limit our ability to learn from natural experiments (and other messy kinds of ecological monitoring). *Ecol. Evol.* <http://dx.doi.org/10.1002/ece3.1782>.
- Dornelas, M. et al., 2014. Assemblage time series reveal biodiversity change but not systematic loss. *Science* 344, 296–299.
- Edwards, D.P., Larsen, T.H., Docherty, T.D.S., Ansell, F.A., Hsu, W.W., Derhe, M.A., Hamer, K.C., Wilcove, D.S., 2011. Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. *Proc. R. Soc. B – Biol. Sci.* 278, 82–90.
- Felton, A., Knight, T.M., Wood, J., Zammit, C., Lindenmayer, D., 2010. A meta-analysis of fauna and flora species richness and abundance in plantations and pasture lands. *Biol. Cons.* 143, 545–554.
- Freligh, L.E., 2002. *Forest Dynamics and Disturbance Regimes: Studies from Temperate Evergreen-Deciduous Forests*. Cambridge University Press, New York.
- Freligh, L.E., Reich, P.B., 2003. Perspectives on development of definitions and values related to old-growth forests. *Environ. Rev.* 11, 9–22.
- Gardner, T.A., Barlow, J., Parry, L.W., Peres, C.A., 2007. Predicting the uncertain future of tropical forest species in a data vacuum. *Biotropica* 39, 25–30.
- Gerstner, K., Moreno-Mateos, D., Gurevitch, J., Beckmann, M., Kambach, S., Jones, H. P., Seppelt, R., 2017. Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Methods Ecol. Evol.* <http://dx.doi.org/10.1111/2041-210X.12758>.
- Ghazoul, J., Burivalova, Z., Garcia-Ulloa, J., King, L.A., 2015. Conceptualizing forest degradation. *Trends Ecol. Evol.* 30, 622–632.
- Gibson, L., Lee, T.M., Koh, L.P., Brook, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J.A., Laurance, W.F., Lovejoy, T.E., Sodhi, N.S., 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* 478, 378–381.
- Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* 4, 379–391.
- Gurevitch, J., Hedges, L.V., 1999. Statistical issues in ecological meta-analyses. *Ecology* 80, 1142–1149.
- Gurevitch, J., Curtis, P.S., Jones, M.H., 2001. Meta-analysis in ecology. *Adv. Ecol. Res.* 32, 199–247.
- Halme, P., Toivanen, T., Honkanen, M., Kotiaho, J.S., Mönkkönen, M., Timonen, J., 2010. Flawed meta-analysis of biodiversity effects of forest management. *Conserv. Biol.* 24, 1154–1156.
- Hargrove, W.W., Pickering, J., 1992. Pseudoreplication: a sine qua non for regional ecology. *Landscape Ecol.* 6, 251–258.
- Harrison, F., 2011. Getting started with meta-analysis. *Methods Ecol. Evol.* <http://dx.doi.org/10.1111/j.2041-210X.2010.00056.x>.
- Hedges, L.V., Olkin, I., 1985. *Statistical Methods for Meta-analysis*. Academic Press, Orlando, Florida.
- Hedges, L.V., Pigott, T.D., 2001. The power of statistical tests in meta-analysis. *Psychol. Methods* 6, 203–217.
- Heffner, R.A., Butler, M.J., Reilly, C.K., 1996. Pseudoreplication revisited. *Ecology* 77, 2558–2562.

- Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.* 54, 187–211.
- Kalies, E.L., Chambers, C.L., Covington, W.W., 2010. Wildlife responses to thinning and burning treatments in southwestern conifer forests: a meta-analysis. *For. Ecol. Manage.* 259, 333–342.
- Kim, R.S., Becker, J., 2010. The degree of dependence between multiple treatment effect sizes. *Multivar. Behav. Res.* 45 (2), 213–238.
- Koricheva, J., Gurevitch, J., 2014. Uses and misuses of meta-analysis in plant ecology. *Ecology* 102, 828–844.
- Koricheva, J., Gurevitch, J., Mengersen, K. (Eds.), 2013. *Handbook of Meta-Analysis in Ecology and Evolution*. Princeton University Press, Princeton and Oxford.
- Lajeunesse, M.J., 2010. Achieving synthesis with meta-analysis by combining and comparing all available studies. *Ecology* 91 (9), 2561–2564.
- Lajeunesse, M.J., 2011. On the meta-analysis of response ratios for studies with correlated and multi-group designs. *Ecology* 92, 2049–2055.
- Lajeunesse, M.J., 2013. Recovering missing or partial data from studies: a survey of conversions and imputations for meta-analysis. In: Koricheva, J., Gurevitch, J., Mengersen, K. (Eds.), *Handbook of Meta-Analysis in Ecology and Evolution*. Princeton University Press, Princeton and Oxford, pp. 195–206.
- Lajeunesse, Marc J., Rosenberg, M., Jennions, M., 2013. Phylogenetic nonindependence and meta-analysis. In: Koricheva, Julia et al. (Eds.), *Handbook of Meta-Analysis in Ecology and Evolution*. Princeton University Press, 284, p. 299.
- Lajeunesse, M.J., 2015. Bias and correction for the log response ratio in ecological meta-analysis. *Ecology* 96 (8), 2056–2063.
- Lau, J., Rothstein, H.R., Stewart, G., 2013. History and progress of meta-analysis. In: Koricheva, Julia, Gurevitch, Jessica, Mengersen, Kerrie (Eds.), *Handbook of Meta-analysis in Ecology and Evolution*. Princeton UP, pp. 407–419.
- Lindenmayer, D.B., Laurance, W.F., 2012. A history of hubris – cautionary lessons in ecologically sustainable forest management. *Biol. Cons.* 151, 11–16.
- Lindenmayer, D.B., Franklin, J.F., Lohmus, A., Baker, S.C., Bauhus, J., Beese, W., Brodie, A., Kiehl, B., Kouki, J., Pastur, G.M., Messier, C., Neyland, M., Palik, B., Sverdrup-Thygeson, A., Volney, J., Wayne, A., Gustafsson, L., 2012. A major shift to the retention approach for forestry can help resolve some global forest sustainability issues. *Conserv. Lett.* 5, 421–431.
- Lortie, C.J., Stewart, G., Rothstein, H., Lau, J., 2013. How to critically read ecological meta-analyses. *Peer J. PrePrints* 1, e38v2.
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell Publishing, Oxford.
- Mayerhofer, M.S., Kernaghan, G., Harper, K.A., 2013. The effects of fungal root endophytes on plant growth: a meta-analysis. *Mycorrhiza* 23, 119.
- Molloy, P.P., Reynolds, J.D., Gage, M.J.G., Mosquera, I., Côté, I.M., 2008. Links between sex change and fish densities in marine protected areas. *Biol. Cons.* 141, 187–197.
- Monserrud, R.A., 2004. Large-scale management experiments in the moist maritime forests of the Pacific Northwest. *Landscape Urban Plan.* 59, 159–180.
- Nakagawa, S., Noble, D.W., Senior, A.M., Lagisz, M., 2017. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. *BMC Biol.* 15 (1), 18. <http://dx.doi.org/10.1186/s12915-017-0357-7>.
- Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Diaz, S., Echeverria-Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Laginha Pinto Correia, D., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. *Nature* 520, 45–50.
- Ogle, K., Barber, J., Sartor, K., 2013. Feedback and modularization in a Bayesian meta-analysis of tree traits affecting forest dynamics. *Bayesian Anal.* 8, 133–168.
- Osenberg, C.W., Sarnelle, O., Cooper, S.D., 1997. Effect size in ecological experiments: the application of biological models in meta-analysis. *Am. Nat.* 150, 798–812.
- Paillet, Y., Berges, L., Hjalten, J., Odor, P., Avon, C., Bernhardt-Romermann, M., Bijlsma, R.J., De Bruyn, L., Fuhr, M., Grandin, U., Kanka, R., Lundin, L., Luque, S., Magura, T., Matesanz, S., Meszaros, I., Sebastia, M.T., Schmidt, W., Standovar, T., Tothmeresz, B., Uotila, A., Valladares, F., Vellak, K., Virtanen, R., 2010. Biodiversity differences between managed and unmanaged forests: meta-analysis of species richness in Europe. *Conserv. Biol.* 24, 101–112.
- Peh, K.S.H., Balmford, A., Bradbury, R.B., Brown, C., Butchart, S.H.M., Hughes, F.M.R., Stattersfield, A., Thomas, D.H.L., Walpole, M., Bayliss, J., Gowing, D., Jones, J.P.G., Lewis, S.L., Mulligan, M., Pandeya, B., Stratford, C., Thompson, J.R., Turner, K., Vira, B., Willcock, S., Birch, J., 2013. TESSA: a toolkit for rapid assessment of ecosystem services. *Ecosyst. Serv.*, 51–57.
- Peterken, G.F., 1974. A method for assessing woodland flora for conservation using indicator species. *Biol. Cons.* 6, 239–245.
- Pullin, A.S., Stewart, G.B., 2006. Guidelines for systematic review in conservation and environmental management. *Conserv. Biol.* 20, 1647–1656.
- Putz, F.E., Zuidema, P.A., Pinard, M.A., Boot, R.G.A., Sayer, J.A., Sheil, D., Sist, P., Elias, Vanclay, J.K., 2008. Improved tropical forest management for carbon retention. *PLoS Biol.* 6, 1368–1369.
- Ramage, B.S., Sheil, D., Salim, H.M.W., Fletcher, C., Mustafa, N.Z.A., Luruthusamay, J. C., Harrison, R.D., Butod, E., Dzulkiply, A.D., Kassim, A.R., Potts, M.D., 2013. Pseudoreplication in tropical forests and the resulting effects on biodiversity conservation. *Conserv. Biol.* 27, 364–372.
- Rosenberg, M., Rothstein, H., 2013. Effect sizes: conventional choices and calculations. In: Gurevitch, J., Koricheva, J., Mengersen, K. (Eds.), *Handbook of Meta-analysis in Ecology and Evolution*. Princeton University Press, 61, p. 71.
- Spake, R., Martin, P.M., Ezard, T.H.G., Newton, A.C., Doncaster, C.P., 2015. A meta-analysis of functional group responses to forest recovery outside of the tropics. *Conserv. Biol.* 29, 1695–1703.
- Spake, R., Van Der Linde, S., Martinez-Suz, L., Bidartondo, M., Newton, A.C., Doncaster, C.P., 2016. Similar biodiversity of ectomycorrhizal fungi in set-aside plantation and ancient broadleaved forests. *Biol. Cons.* 194, 71–79.
- Stewart, G., 2010. Meta-analysis in applied ecology. *Biol. Lett.* 6, 78–81.
- Vanderwell, M.C., Malcolm, J.R., Mills, S.C., 2007. A meta-analysis of bird responses to uniform partial harvesting across North America. *Conserv. Biol.* 21 (5), 1230–1240.
- Veroniki, A.A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J.P.T., Langan, D., Salanti, G., 2016. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res. Syn. Meth.* 7, 55–79.
- Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48.
- Whittaker, R.J., 2010. Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. *Ecology* 91, 2522–2533.
- Wirth, C., Messier, C., Bergeron, Y., Frank, D., Frankhänel, A., 2009. Old-growth forest definitions: a pragmatic view. In: Wirth, C., Gleixner, G., Heimann, M. (Eds.), *Old Growth Forests: Function, Fate, and Value*. Springer, Berlin, pp. 11–33.
- Woodcock, P., Edwards, D.P., Fayle, T.M., Newton, R.J., Khen, C.V., Bottrell, S.H., Hamer, K.C., 2011. The conservation value of South East Asia's highly degraded forests: evidence from leaf-litter ants. *Proc. R. Soc. B* 366, 3256–3264.