Meta-analysis in Forest Biodiversity-Ecosystem Functioning Research

Dissertation

zur Erlangung des
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der
Naturwissenschaftlichen Fakultät I - Biowissenschaften -

der Martin-Luther-Universität
Halle-Wittenberg,

vorgelegt

von **Herrn Stephan Kambach (Dipl.-Biol.)**
geb. am 21.05.1987 in Leipzig

Datum der Verteidigung: 03. Juli 2018

Gutachter:
Prof. Dr. rer. nat. Helge Bruelheide
Institut für Biologie, Martin-Luther-Universität Halle-Wittenberg
Am Kirchtor 1, 06108 Halle, Deutschland
Tel. +49 345 55 26222
helge.bruelheide@botanik.uni-halle.de

Prof. Dr. rer. nat. Ingolf Kühn
Institut für Biologie, Martin-Luther-Universität Halle-Wittenberg, sowie
Helmholtz-Zentrum für Umweltforschung
Theodor-Lieser-Straße 4, 06120 Halle, Deutschland
Tel. +49 345 558-5512
ingolf.kuehn@ufz.de

Prof. PhD Messier Christian
Isfort (Institut Des Science De La Forêt Tempérée)
58 Rue Principale, Ripon, QC J0V 1V0, Kanada
Tel. +1 819-595-3900 ext. 2904
ch.messier@gmail.com

What will be will be.

Every river flows into the sea.

But it's never enough.

Architects - Doomsday

Epitaph Records, 2017

Copyright notice

Table of Contents

1 | Summary

Meta-analysis refers to a toolbox for conducting systematic quantitative reviews in order to derive unbiased conclusions on the consistency and magnitude of effects across a body of published literature. Although meta-analyses have yet rarely been applied to summarize relationships between forest biodiversity and ecosystem functioning (BEF relationships), they were already used to confirm the generally positive effects of increasing tree diversity on forest productivity and resistance against insect herbivory. Whether and how heterogeneity and even contradictive findings between different studies can be explained with the differences in the applied study design (observational versus experimental studies), different locations and climatic conditions or differences in the tree species composition is, however, less well understood. In my thesis, I thus applied meta-analytical tools in order to investigate how these differences in the study design could moderate the shape of the relationships between tree diversity and, first, tree growth and, second, forest resistance to insect herbivory.

In the first study of this thesis (chapter 5), I performed an extensive literature search and summarized published relationships between the Shannon diversity of tree species and different aspects of insect herbivory, namely herbivore abundance, damage inflicted, incidence rate and species richness. I found the mean effect of tree diversity on insect herbivory to be insignificant and independent from the study design (observational versus experimental studies). Along a global gradient of increasing mean annual temperature, however, I found the relationship between tree diversity and the damage inflicted to become more negative (i.e. shifting toward associational resistance) whereas the relationship with the incidence rate became more positive (i.e. shifting towards associational susceptibility) with increasing mean annual temperature. These results provided the first evidence that associational effects in forest ecosystems might shift along global gradients, an effect that could partly account for diverging results of studies that are conducted in different locations.

In the second study, I tested whether the responses of up to 64 different tree species to species mixing can be transferred between three different research approaches (five national forest inventories, 209 comparative plots in six established forests and six tree diversity experiments). The data were compiled and harmonized within the European FunDivEUROPE project. Although

those approaches were conducted in partly overlapping climatic conditions and species compositions, tree species' responses to mixing could generally not be transferred between the three research approaches. In accordance with a recently published meta-analysis, I found that the effects of species mixing on species growth were positive and stronger in observational research approaches (i.e. in national forest inventories and comparative forest plots) than in tree diversity experiments for which I found the effects of species mixing to be not (yet) significant. Since none of the investigated tree species showed a generally negative response to species mixing, I concluded that many, if not most, investigated monospecific tree communities could be diversified without negative or even with positive effects on tree growth.

Although the first two studies of this thesis could be based on complete data sets, incompletely reported or missing data (especially in the case of effect sizes variances) is still a frequent problem in ecological meta-analyses. In the literature it is generally recommended to impute those missing variances and there exists a range of easily applicable imputation techniques. Still, there is no comprehensive overview on the different imputation techniques and their reliability under correlations structures that might be hidden in meta-analytical data sets. In the third study of this thesis, I therefore simulated meta-analysis data sets with missing variance estimates and different correlation structures and tested the performance of ten different methods that can be applied to treat those missing values. My results clearly showed that the omission of incompletely reported studies, which is a frequently applied practice in published meta-analyses, can lead to biased summary effect sizes. Imputations via linear regression and predictive mean matching were particularly effective in accommodating hidden correlation structures. These findings should encourage researchers to routinely impute missing variance estimates and could thereby raise the reliability and precision of future meta-analysis results, not only in the ecological sciences.

In summary, my contributions to the emergent field of meta-analysis in forest BEF relationships encompassed the findings that: (i) tree diversity effects on insect herbivory could change along a not yet considered global gradient, (ii) species-specific effects of tree diversity on growth cannot (yet) be transferred between different research approaches and (iii) imputation of missing variance estimates in meta-analysis dataset can increase the reliability and precision of the obtained grand mean effect (under consideration of different correlation structures).

2 | Zusammenfassung

Unter dem Begriff der ˈMetaanalyseˈ werden Methoden systematischer und quantitativer Reviews zusammengefasst, welche objektive Aussagen zur Größe und Konsistenz eines untersuchten Effektes über die veröffentlichte Literatur zu ermöglichen. Obwohl Metaanalysen noch selten angewandt wurden um veröffentlichte Zusammenhänge zwischen der Biodiversität und Funktion von Wald-Ökosystem zusammenzufassen, konnten damit schon die generell positiven Effekte einer höherer Baum-Diversität auf die Waldproduktivität und Resistenz gegen Insektenherbivoren nachgewiesen werden. Weniger gut untersucht ist, ob und wie scheinbar widersprüchliche Ergebnisse verschiedener Studien mit deren Unterschieden im Studiendesign (Beobachtungs- oder Experimentalstudien), verschiedenen Standorten und klimatischen Bedingungen oder verschiedenen Artgemeinschaften erklärt werden könnten. In dieser Dissertation habe ich mittels meta-analytischer Methoden untersucht, wie die Beziehung zwischen Baum-Diversität und Wachstum sowie Resistenz gegenüber Insektenherbivoren durch die genannten Unterschiede im Studiendesign erklärt werden könnten.

In Studie Nummer Eins (Kapitel 5) habe ich, nach umfangreicher Literaturrecherche, den Zusammenhang zwischen der Shannon-Diversität der Baumgemeinschaft und deren Belastung durch Insektenherbivoren untersucht, wobei ich dabei unterschieden habe ob es sich um die Abundanz, den Schadens, die Inzidenzrate oder die Artenzahl der Herbivoren handelte. Ich fand heraus, dass der generelle Effekt der Baumdiversität auf die Insektenherbivoren nicht signifikant war und sich auch nicht zwischen Beobachtungs- und Experimentalstudien unterschied. Mit steigender mittlerer Jahrestemperatur zeigte sich jedoch ein zunehmend negativer Zusammenhang zwischen Baumdiversität und Herbivorenschaden (d.h. eine erhöhte Resistenz mit steigender Baumdiversität) und ein zunehmend positiver Zusammenhang mit der Inzidenzrate (d.h. eine erhöhte Anfälligkeit mit steigender Baumdiversität). Mit diesen Ergebnissen konnte ich einen ersten Erweis bringen, dass sich Diversitätseffekte in Waldökosystemen entlang eines globalen Gradienten ändern können. Damit ließen sich auch teilweise widersprüchliche Ergebnisse zwischen Studien, welche an verschiedenen Orten durchgeführt wurden, erklären.

In der der zweiten Studie dieser Dissertation habe ich das Wachstum von bis zu 64 Baumarten in Mischung und Monokulturen verglichen. Ziel dieser Analyse war, zu testen, ob die beobachteten Mischungseffekte zwischen drei verschiedenen Forschungsansätzen übertragen werden können (und zwar zwischen fünf nationalen Waldinventuren, 209 Plots in sechs naturnahen Waldtypen und sechs Baumdiversitätsexperimenten). Die Daten für meine Analysen wurden im Rahmen des europäischen FunDivEUROPE-Projektes bereitgestellt und harmonisiert. Obwohl die drei untersuchten Forschungsansätze in teilweise überlappenden klimatischen Bedingungen und mit ähnlichen Artgemeinschaften durchgeführt wurden, konnten die artspezifische Mischungseffekte nicht übertragen werden. Übereinstimmend mit einer kürzlich veröffentlichten Metaanalyse konnte ich weiterhin zeigen, dass die untersuchten Diversitätseffekte in Beobachtungsstudien; also in den nationalen Waldinventuren und naturnahen Wäldern; signifikant positiv waren wohingegen in Experimentalstudien; also in den Baumdiversitätsexperimenten; (noch) keine signifikanten Diversitätseffekte zu beobachten waren. Da keine der untersuchten Baumarten ein generell vermindertes Wachstum in Artmischung zeigte, kam ich in dieser Studie zu dem Schluss, dass viele, wenn nicht fast alle der untersuchten Monokulturen diversifiziert werden könnten, ohne negative und teilweise sogar mit positiven Effekten auf deren Wachstum.

In der ersten und zweiten Studie dieser Dissertation konnte ich meine Untersuchungen auf vollständige Datensätze stützen. Unvollständige oder schlicht fehlende Daten (besonders solcher, welche die Varianz von Effektgrößen betreffen) sind jedoch noch oft Problem in ökologischen Metaanalysen. Zwar wird in der Literatur empfohlen, diese fehlenden Varianzwerte zu schätzen (i.e. zu imputieren) und es werden auch verschiedene Methoden zu Varianzschätzung beschrieben. Allerdings fehlte bisher ein umfassender Überblick über die verschiedenen Schätzungsmethoden, sowie die Auswirkungen von verschiedenen Korrelationsstrukturen, welche in Metaanalyse-Datensätzen auftreten könnten, auf die imputierten Werte. In der dritten Studie dieser Dissertation habe ich daher anhand simulierter Datensätze mit verschiedenen Korrelationsstrukturen und fehlenden Varianzwerten getestet, inwieweit zehn verschiedenen Verfahren mit diesen fehlenden Werten umzugehen, das Ergebnis von Metaanalysen beeinflussen können. Dabei zeigte sich deutlich, dass die Nicht-Einbeziehung von Studien mit fehlenden Varianzwerten, wie sie in publizierten Metaanalysen oft angewendet wurde, zu verfälschten Ergebnissen führen kann. Für mögliche Korrelationen in Metaanalyse-Datensätzen konnten insbesondere die Schätzung mittels linearer Regression oder *predictive mean matching* gut korrigieren. Mit diesen Ergebnissen konnte ich zeigen, dass fehlende Varianzwerte standardmäßig geschätzt werden sollten, wodurch sich dann die Glaubwürdigkeit und Präzision von zukünftiger Metaanalyse-Ergebnissen, nicht nur in der Ökologie, erhöhen würde.

Mit meiner Dissertation trage ich, zu dem sich noch entwickelndem Gebiet der Metaanalyse von Zusammenhängen zwischen Wald-Biodiversität und Ökosystemfunktionen, die folgenden Hauptergebnisse bei: (i) Effekte von der Baumdiversität auf Insektenherbivoren könnten sich entlang eines, noch nicht berücksichtigten, globalen Gradienten ändern, (ii) artspezifische Mischungseffekte auf das Wachstum verschiedener Baumarten können (noch) nicht zwischen verschiedenen Forschungsansätzen übertragen werden und (iii) unter Berücksichtigung möglicher Korrelationsstrukturen kann die Schätzung fehlender Varianzwerte die Glaubwürdigkeit und Präzision von Metaanalyse-Ergebnissen erhöhen.

3 | Introduction

3.1 | Research Synthesis and Meta-analyses

The accumulation of scientific knowledge is an iterative process in which previous findings are reviewed to derive new hypotheses, which are then tested and whose results lead to the development of new hypotheses (Ford 2000). A prerequisite for the functioning of such a stepwise extension of knowledge is that previous findings must be consistent and uncertainties known. In theory, two studies that investigate the same phenomenon with a similar study-design should yield comparable results on both, the direction and the magnitude of the investigated relationships. Given that all sources of heterogeneity are controlled for, significant findings as well as error probabilities should be reproducible in subsequent replication studies.

However, a recent survey across 1,576 scientists revealed that more than 70 percent had already tried and failed to reproduce another scientist’s results. More than half of the surveyed scientists had even failed to reproduce some of their own results (Baker 2016). In the field of psychology, the Open Science Collaboration could only replicate 39 out of 100 peer-reviewed studies that were all published in high-ranked journals (Open Science Collaboration 2015). In the field of medicine, Freedman *et al.* (2015) even concluded that 50 percent of all US preclinical studies could suffer from irreproducibility. Based on these numbers, the US alone might spend 28 billion US$ each year on medical studies whose results could not be replicated later on (Freedman *et al.* 2015).

Ecological relationships, especially if observed in natural ecosystems, can be affected by a plethora of factors that researchers cannot control for or might even not be aware of. Thus, it is very likely that ecological studies suffer from a similar problem of irreproducibility as other research domains although this issue has not yet been vigorously addressed because the exact replication of ecological experiment is often difficult if not impossible (Fidler *et al.* 2017).

Failed attempts to replicate significant results led to the proclamation of a general “replication crisis”. In the view of the public media, the scientific approach to knowledge generation and the scientific community as a whole suffered from a loss of trustworthiness, as reflected by headlines like: “Many psychology findings not as strong as claimed, study says" (The New York Times 2015), “Science falling victim to 'crisis of narcissism” (The Guardian 2017), “Most scientists 'can't replicate studies by their peers” (BBC News 2017). In the scientific community, this critique has triggered extensive debates on the misapplication of p-values and effect size significances (e.g. Nuzzo 2014, Greenland *et al.* 2016, Wasserstein & Lazar 2016, Amrhein & Greenland 2017, Benjamin *et al.* 2017) as well as deficits in the scientific processes of data-analysis, presentation, peer-review and publication (e.g. Lindsay 2015, Morey *et al.* 2016, Lee & Moher 2017, Martin & Clarke 2017, Naik 2017).

Meta-research (or research synthesis) approaches have prominently been advocated as one potential measure to solve the replication crisis (Schooler 2014) and to evaluate research evidence and reproducibility in ecological and evolutionary sciences (Fidler *et al.* 2017). The term research synthesis refers to the scientific attempt of integrating multiple findings on a specific research question across different sources of primary research in order to find generalizations and resolve conflicts between contradicting study results (definition adapted from Koricheva *et al.* 2013). Syntheses can be achieved either via qualitative and semi-quantitative review narratives or via quantitative statistical analyses across published results and data sets. In 1976, Gene V. Glass coined the “meta-analysis” which he described as “the analysis of analyses", i.e. "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass 1976, p. 3). Today, the term meta-analysis usually refers to a toolbox of systematic and reproducible steps that encompass the search for literature, the process of data extraction, the statistical aggregation of study outcomes (effect sizes) and the detection and correction of any bias in the collated literature body.

The very first meta-analyses were conducted in the medical and the social sciences (by e.g. Pearson 1904, who already applied key methods of modern meta-analyses, and Chalmers *et al.* 1977). Currently, quantitative meta-analyses and systematic reviews are regarded as fundamental methods for "...integrating individual clinical expertise with the best available external clinical evidence from systematic research" in order to support an evidence-based medicine paradigm that guarantees the "...use of current best evidence in making decisions about the care of individual patients..." (Sackett *et al.* 1996, p. 71). The further improvement of meta-analytical standards is nowadays pushed by global initiatives like the Cochrane network that is dedicated towards an “improved health where decisions about health and health care are informed by high-quality, relevant and up-to-date synthesized research evidence” (Our

Vision, www.cochrane.org) and the Society for Research Synthesis and Methods whose objective is to "promote and encourage the use of appropriate and robust methods of synthesis" (www.srsm.org). In recent years, the application and development of meta-analytical methods in ecological research were fostered by, for instance, the publication of the Handbook of Meta-analysis in Ecology and Evolution (Koricheva *et al.* 2013), a special issue in the Journal of Ecology (*Meta-Analysis in Plant Ecology*, Gómez-Aparicio & Lortie 2014) and through various methodological articles in the journals *Methods in Ecology and Evolution* and *Research Synthesis Methods*.

Two precedent conditions should be met to obtain reliable research syntheses. All research findings should be published regardless of their statistical significance and all results should be reported completely, i.e. with error margins, variance estimates, numbers of observations and additional measures necessary to evaluate the precision of the obtained results. Unfortunately, missing and incompletely reported data are a prevailing problem in ecological studies (Parker *et al.* 2016). Nearly half of the ecological meta-analyses published until 2015 had to deal with incompletely reported results (see chapter 7, study 3 of this thesis). Recent methodological and computational advances now enable researchers to impute those missing data in order to assess and increase the reliability of the synthesized effect sizes. However, we are still missing a comprehensive overview on the imputation techniques that can be applied and how hidden correlation structures in meta-analysis datasets could affect those imputations. Such an overview could foster the future application of adequate imputation techniques, especially in research fields to which meta-analytical tools had only recently been introduced, like in the case of forest biodiversity-ecosystem functioning relationships.

3.2 | Tree Diversity and Forest Functioning

Human well-being depends to a large degree on functions and services provided by natural ecosystems, be it the basic provision of food and crafting material, the safeguarding of future supplies and liveable conditions or the provision of recreational spaces and cultural identification (de Groot *et al.* 2010; Costanza et al. 1997; Reid *et al.* 2005). Every ecosystem, and earth as a whole, is inhabited by a plethora of interacting and interdependent species (with planetary estimates ranging between one and six million species, Larsen *et al.* 2017). With the oldest fossils dating back roughly 350,000 years (Hublin *et al.* 2017), the modern human species can still be regarded as a quite new member of the planetary universe of ecosystems, species and genes which altogether determine the biodiversity of all living things on earth. Human imprints on the global atmosphere and thus ecosphere can be traced back to as early as 8,000 years, more specifically to the onset of agricultural and deforestation practices (Ruddiman 2003). From then on, humans impacted or transformed nearly every ecosystem on earth up to the point that scientists recently proclaimed the start of a new geologic time epoch, the so-called “Anthropocene” as characterized by unprecedented changes in global atmospheric and stratigraphic processes (Waters *et al.* 2016).

As a result of those unprecedented changes, many species are threatened by decreases in their global number of individuals, populations and a decreasing amount of suitable habitats. This human-induced global annihilation of biodiversity (Ceballos *et al.* 2017) eventually threatens the survival of so many species that some scientists proclaimed the beginning of the sixth wave of mass extinctions (Barnosky *et al.* 2011, Ceballos *et al.* 2015, Ceballos *et al.* 2017) as global extinction rates reached alarmingly unprecedented peak highs (Pereira *et al.* 2010, Barnosky *et al.* 2011, Carlson *et al.* 2017).

In most ecosystems, a loss of biodiversity, regardless if it is species, genetic, or functional diversity, often leads to a decrease in ecosystem functioning and potentially also ecosystem services, i.e. those functions that are beneficial to human well-being (Cardinale *et al.* 2002, Duraiappah *et al.* 2005, Hooper *et al.* 2005, Díaz *et al.* 2006, Cardinale *et al.* 20012, Hooper *et al.* 2012, Naeem *et al.* 2012, Turnbull *et al.* 2016). Exemplary ecosystem services that depend on the biodiversity of ecosystems are the production of biomass, the cycling of nutrients, the systems resistance against pests and disturbance threats, the purification of toxic substances and the potential to attenuate the drivers and consequences of global climate and land-use changes (Cardinale *et al.* 2012). To what extent ecosystem functioning depends on the biodiversity of the investigated system (biodiversity-ecosystem functioning relationships, hereafter BEF relationships) and which are the drivers of those BEF relationships are still some of the central topics in the fundamental and applied fields of ecological research.

Up to date, the most comprehensive evidence for prevailing positive BEF relationships in terrestrial ecosystems stems from experimental grassland communities. A loss in primary producer diversity has been shown to decrease, for instance, the productivity (Naeem *et al.* 1994) and resistance to major drought events, invasion by exotic species and infections with fungal diseases in experimental grassland ecosystems (Tilman & Downing 1994, Knops *et al.* 2007). The magnitude of those grassland BEF relationships increases over time (e.g. Eisenhauer *et al.* 2012, Reich *et al.* 2012, Ravenek *et al.* 2014, Meyer *et al.* 2016).

Globally, grassland ecosystems cover between 20 and 40 percent of the terrestrial surface and thus a higher proportion than forest ecosystems, which cover between 28 and 31 percent (Figure 3.1, Hansen *et al.* 2013, Latham *et al.* 2014, FAO statistical data from 2008 - www.fao.org). However, the majority of the global terrestrial biodiversity (between 50 and 80 percent) can found in those forest ecosystems (CBD 2010, IUCN 2012) and the approximately 3 trillion trees that are growing on planet earth belong to more than 60,000 species (Crowther *et al.* 2015, Beech *et al.* 2017). Today, these forests and their inhabiting biodiversity face severe threats. Between 2000 and 2012, the global extent of forest ecosystems decreased by approximately 5.1 to 6.1 percent, with subtropical forests experiencing the highest reduction (between 7.8 and 11.3 percent, Hansen *et al.* 2013). Since tropical and subtropical forests also harbour the highest number of tree species (Figure 3.2), the ongoing destruction of these forests likely puts many tree species under a severe risk of extinction. Accordingly, the recently published global database of tree species (*GlobalTreeSearch*, Beech *et al.* 2017) shows that at least 9,300 tree species face severe extinction threats whereas ter Steege *et al.* (2015) estimated that most of the more than 40,000 tropically distributed tree species might already qualify as being globally threatened by extinction.

Losing tree species, analogue to losing grassland species, affects forests functioning (Nadrowski *et al.* 2010, Liang *et al.* 2016, Brockerhoff *et al.* 2017). Studying the effects of tree diversity on forest functioning is, however, more laborious than studying the effects of grassland diversity. Forests harbour a complex structural stratification and the longer life-cycle of trees conditions a longer time span for forests to mature and shape interaction between the different tree species and the dependent trophic levels (Scherer-Lorenzen 2014). Still, losing tree species has already been found to decrease the productivity of forests (Piotto 2008, Zhang *et al.* 2012, Chisholm *et al.* 2013, Liang *et al.* 2016), decrease their resistance to drought events (Jucker *et al.* 2016), increase the herbivore pressure by specialised insects (Jactel & Brockerhoff 2007) and generally decrease forest multifunctionality (i.e. the ability to simultaneously provide high levels of multiple functions, van der Plas *et al.* 2016). In its 2013 EU Forest Strategy the European Union explicitly embraced the target to develop forest management practices that favour sustainable and multifunctional forests (European Commission 2013).

Many, if not all, of forest BEF relationships are context-sensitive, meaning that their magnitude and direction depend on the composition and ontogeny of the tree species, the environmental context and the extent of the investigated forest (e.g. Ratcliffe *et al.* 2017, Chisholm *et al.* 2013). Diverging and sometimes even contradicting findings on forest BEF relationships could thus partly be explained by differences in the applied study design (e.g. Madrigal-González *et al.* 2016, Ratcliffe *et al.* 2017). Here, meta-analytical tools can be applied to test whether diverging effect of tree diversity (either between studies or between research sites) can be (post-hoc) related to certain differences in the study design or site conditions and thereby contribute to the development of new hypotheses on the drivers of forest BEF relationships.

3.3 | Meta-analysis in Forest BEF Research

In order to depict ecosystem management options for forest holders and policy makers which conjointly safeguard human well-being and natural biodiversity (Adams *et al.* 2004) the value of biodiversity for forest functioning must be assessed in a systematic manner (Díaz *et al.* 2015). Generalizations across the large and still growing body of literature on forest BEF relationships (Figure 3.3) can be drawn from either qualitative and semi-quantitative reviews or from meta-analyses, i.e. from replicable quantitative summaries of published research results. Especially the aspects of reproducibility and context-dependency can often only be tackled via research synthesis (Fidler *et al.* 2017) since individual BEF studies can often only be methodologically and spatially restricted.

In ecology, the first meta-analyses were published in 1991 (Jarvinen 1991, Marchant & McGrew 1991). The first study to mention forest biodiversity, ecosystem functioning and meta-analysis, according to the Web of Science core collection, was published in 2001 (see Figure 3.3). This study already demonstrated the value of meta-analytical tools for forest conservation by summarizing the effects of even- versus uneven-aged forestry on different species groups

(Gram *et al.* 2001). The first meta-analysis that summarized the results of previous forest BEF studies (according to the Web of Science core collection) was published in 2007 and tested the effects of tree diversity on insect herbivory (Jactel & Brockerhoff 2007). This study also showed how different study designs, for instance whether specialist or generalist herbivores were studied, can impact the obtained effects of tree diversity. Since then, only two to six studies that that mentioned forest BEF and meta-analysis were published every year. These numbers increased only as recently as 2016, when 14 studies were published (see Figure 3.3). Of those 62 Web of Science-listed publications that mentioned meta-analysis and forest BEF relationships, only three studies actually conducted a meta-analysis on the effects of tree diversity on forest functioning (see Table 3.1), more precisely on insect herbivores (Jactel & Brockerhoff 2007) and tree productivity (Lewandowska *et al.* 2016, Dieler *et al.* 2017). Both forest functions are of high interest for forest owners that strive for a high and stable production of wood.

Previously published meta-analyses already confirmed that in forest ecosystems a higher richness of tree species generally increases the productivity of trees (e.g. Piotto 2008, Zhang *et al.* 2012, Chisholm *et al.* 2013, Liang *et al.* 2016) and reduces the pressure from insect herbivores (Jactel & Brockerhoff 2007, Castagneyrol *et al.* 2014). Although the generality of a positive effect of tree species diversity on those two functions might thus be regarded as established in the scientific community, individual studies still obtained largely divergent results. Whether these differences can be related to differences in the study design (observational versus experimental), study location or tree species composition is less clear but could also be investigated with (post-hoc) meta-analytical approaches.

Table 3.1 | Studies collated in the Web of Science Core collection that were found with the search term *Topic: (forest meta-analysis)* with an indication if the respective study focussed on the relationship between tree biodiversity and forest functioning.

Study	Topic	Focus on tree BEF?
Gray *et al.* 2007	Effects of forest disturbance on bird assemblages	No
Jactel & Brockerhoff 2007	**Effects of tree species diversity in insect herbivory**	**Yes**
Koellner & Scholz 2008	Summarize the species richness and ecosystem damage potential for 53 land use types	No
Attwood *et al.* 2008	Effects of land use intensification on arthropod assemblages	No
Philpott *et al.* 2008	Effects of coffee management on ant, bird and tree assemblages	No
Myers & Harms 2009	Effects of seed supply on the plant assemblages	No
Bénitez-López *et al.* 2010	Effects of infrastructure on mammal and bird assemblages	No
Forsman *et al.* 2010	Effects of small scale disturbances on bird assemblages	No
Ojea *et al.* 2010	Effects of forest type and diversity on human welfare	No
Riffell *et al.* 2011	Effects of coarse wood debris on species diversity	No
Hill *et al.* 2011	Effects of forest fragmentation on bird assemblages	No
Fontaine & Kennedy 2012	Effects of forest fire on mammal and bird assemblages	No
Del Toro *et al.* 2012	Effects of ant and rodents on seed movement	No
De Beenhouwer *et al.* 2013	Effects of coffee and cacao agroforestry on biodiversity and ecosystem services	No
Vellend *et al.* 2013	Summarize temporal changes in plant assemblages	No
Wolowski *et al.* 2014	Summarize the incidence of pollen limitation	No
Nekola and McGill 2014	Scale dependency of the distance decay relationship	No
Vellend *et al.* 2014	The correlation between species and genetic diversity	No
Garciá-Palacios *et al.* 2015	Effects of global change drivers on soil communities and functions	No
Blanco *et al.* 2015	Summarize the forest owner/manager typologies and decision mechanisms	No
Seibold *et al.* 2015	Effects of dead wood on species diversity	No
Spake *et al.* 2015	Effects of forest recovery on fungi, lichens, and beetles assemblages	No
Shantz *et al.* 2016	Effects of nutrients enrichment on phototrophs and heterotrophs	No
Byrom *et al.* 2016	Summarize the effectiveness of pest control	No
Zaitsev *et al.* 2016	The effect time since last fire on forest soil fauna	No
Costantini *et al.* 2016	Effects of logging on forest mammal and bird assemblages	No
Anderegg *et al.* 2016	Summarize the relationship between tree traits and mortality from drought	No
Lewandowska *et al.* 2016	**The effects of resource availability on BEF relationships**	**Yes**
Trivedi *et al.* 2016	Effects of agricultural management on microbial communities and properties	No
Neuschulz *et al.* 2016	Effects of forest disturbance on plant regeneration processes	No
Fuzessy *et al.* 2016	Effects of primate gut passage on seed germination	No
Dove & Hart 2017	Effects of fire on fungi assemblages	No
Bukoski *et al.* 2017	Summarize carbon stock estimates in mangrove forests	No
Rossetti *et al.* 2017	Effects of habitat fragmentation on insect herbivores	No
Hevia *et al.* 2017	Summarize the link between drivers of change and ecosystem services via functional traits of	No
Magura *et al.* 2017	Effects of edge history on beetle assemblages	No
Craven *et al.* 2017	Effects of earthworm introduction on plant assemblages	No
Martin *et al.* 2017	Effects of traits and climate on the impact on invasive plants on carbon pools	No
Westgate *et al.* 2017	Summarize relationships in the community structure of different taxonomic groups	No
Willms *et al.* 2017	Effects of thinning and burning on plant assemblages	No
Dieler *et al.* 2017	**Effects of forest management on tree diversity and productivity**	**Yes**

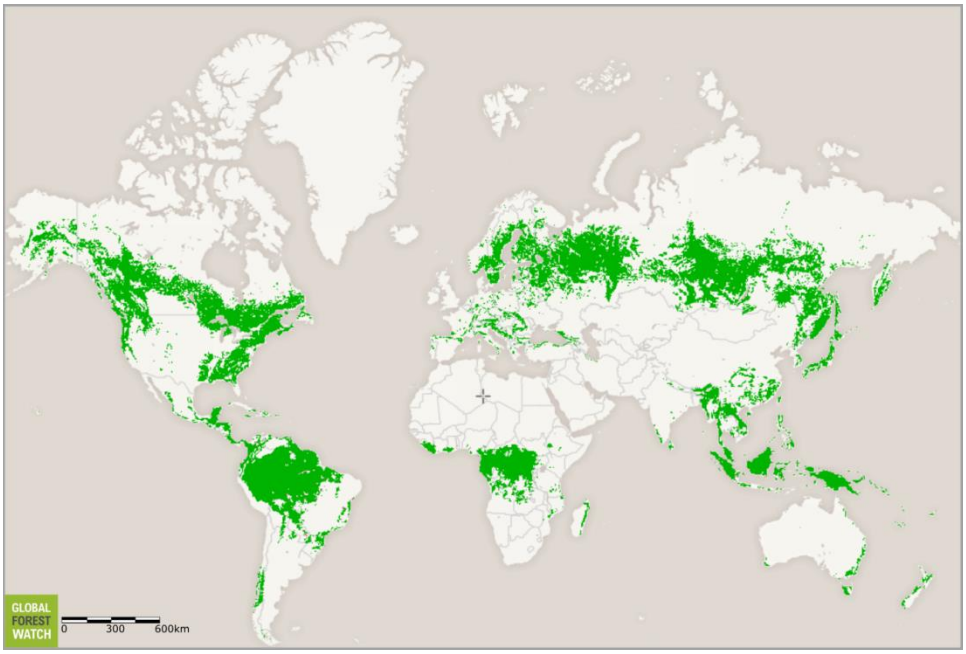


Figure 3.1 | Global forest map highlighting areas with a tree crown density of more than 50 percent (per 30m x 30m pixel) in the year 2000. The map was created with the interactive Global Forest Watch MAPBUILDER. Source: UMD, Google, USGS, NASA and Hansen *et al.* (2013).

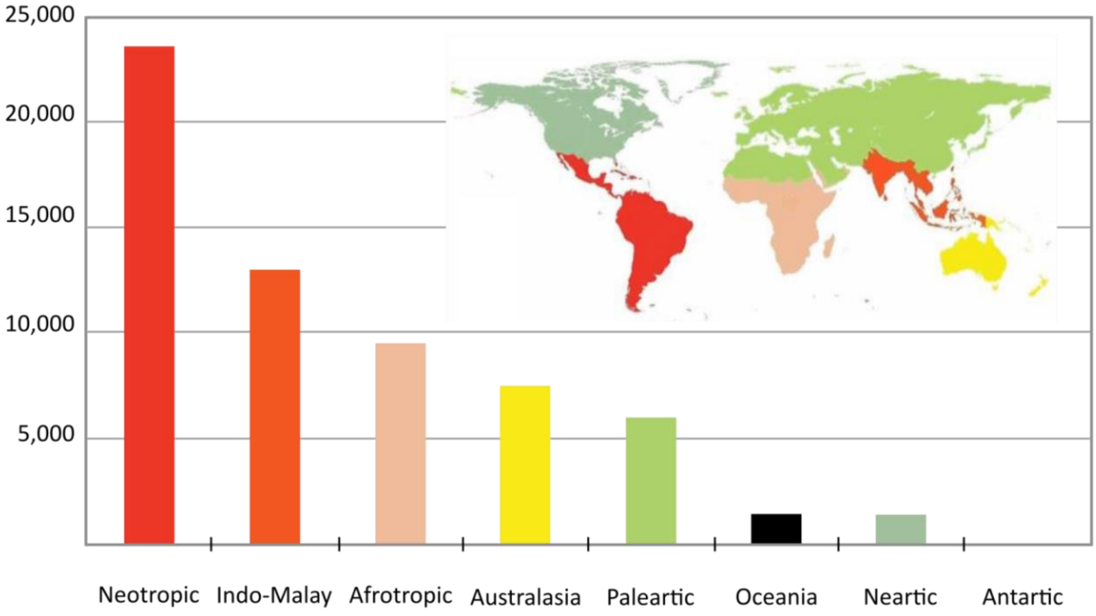


Figure 3.2 | Number of tree species per global biome. The figure is adapted from Figure 5 in the original publication of Beech *et al.* (2017).

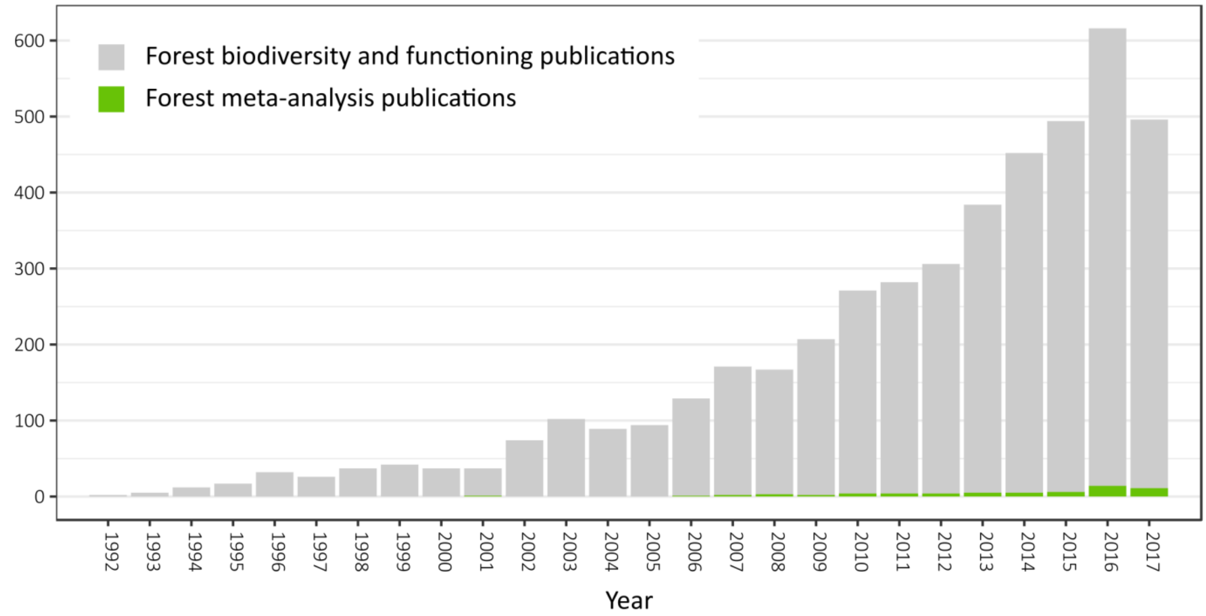


Figure 3.3 | Number of publications per year as indexed in the Web of Science core collection. Grey bars represent those publications that were found with the search term *TOPIC: (biodivers* forest function*)*. Green bars represent those publications that were found with the search term *TOPIC: (forest meta-analysis)*. Both searches were conducted at 2 pm, 6th October 2017.

4 | Scope of this Thesis

The aim of this thesis is to expand the understanding on how the effects of tree diversity on forest functioning can be mediated by the study design applied. Differences in study design, in this thesis, comprise the differentiation between experimental and observational studies, as well as differences in the study location and tree species composition. Those differences in the study design can rarely be investigated within individual studies. I thereby focussed on the effects of tree diversity on two important forest functions, namely tree growth and the resistance to insect herbivores.

In study 1 of this thesis, I asked whether the effects of species mixing on tree growth are consistent between different research approaches (experiments and forest observations) and between different forest types within Europe. More specifically I tested whether species' individual responses to mixing can be transferred between three distinct research approaches (tree diversity experiments, national forest inventories and comparative plots in established forests). All three research approaches were applied across the European continent and partly overlapped in their forest types, climatic conditions and species compositions. This extensive analysis was only possible due to collaborations within the multi-national FunDivEUROPE research network. I thereby aimed at identifying those tree species and forest types that generally benefit from mixed as compared to monospecific tree communities. Furthermore, I provide the currently most extensive test on whether tree diversity effects that are obtained with experimental study designs can be transferred to observational studies and thus be interpolated up to national forest inventories.

In study 2 of this thesis, I demonstrated how the location of the study and the studied aspect of insect herbivory can determine the observed relationship between tree diversity and insect herbivory. I therefore compiled a literature database for which I extracted the correlation coefficient between the Shannon diversity of tree species and the investigated aspect of insect herbivory (damage, abundance, incidence rate and species richness). To my knowledge, this has been the first study to test if associational effects of tree species could change along a gradient from boreal to tropical studies.

In study 3 of this thesis, I then focus on methodological issues of incompletely reported study results and their effects on meta-analysis results. During a collaborative meta-analysis project, we recognized that a large proportion of the compiled studies did not report the variance measures that would be necessary to assess the precision and thus the weight that studies should get in a meta-analysis. We eventually decided to foster the adoption of higher publication standards in future ecological studies by publishing a list of core issues and reporting standards that should facilitate the inclusion, citation and thus outreach of primary research in comprehensive meta-analyses (S4.1, Gerstner *et al.* 2017). Although from the literature it is evident that missing variance values can and should be imputed, it is largely unclear which imputation technique should be applied under the assumption of hidden structures in the literature dataset. To get an idea on the frequency and severity of missing data in ecological meta-analyses, I start this study with an overview on how many ecological meta-analyses actually encountered missing data and how they treated those incompletely reported studies. I then tested whether different techniques can be applied to impute missing standard deviations or sample sizes in artificial meta-analysis data sets and whether hidden dependencies in the data sets can be accommodated or will lead to biased imputations and thus meta-analysis results. The results of the literature review and the simulation study should, in conjunction, help to raise researchers' awareness of the consequences that missing values can have for the results of subsequent meta-analyses. By showing that the imputation of missing values can in fact increase the trustworthiness of a meta-analysis, I intent to increase the quality and reliability of future meta-analyses; also outside of the field of ecology.

5 | Study 1 - How Do Trees Respond to Species Mixing in Experimental Compared to Observational Studies?

Stephan Kambach*[1,2,3], Eric Allan[4,5], Simon Bilodeau-Gauthier[6], David A. Coomes[7] Josephine Haase[8,9], Tommaso Jucker[7,10], Georges Kunstler[11], Sandra Müller[8], Charles Nock[8], Alain Paquette[6], Fons van der Plas[12,13], Sophia Ratcliffe[13], Fabian Roger[14], Paloma Ruiz-Benito[15], Michael Scherer-Lorenzen[8], Harald Auge[2,3], Olivier Bouriaud[16], Bastien Castagneyrol[17], Jonas Dahlgren[18], Lars Gamfeldt[14], Hervé Jactel[19], Julia Koricheva[20], Gerald Kändler[21], Aleksi Lehtonen[22], Bart Muys[23], Quentin Ponette[24], Nuri Setiawan[25], Thomas Van de Peer[23,25], Kris Verheyen[23,25], Miguel A. Zavala[15], Helge Bruelheide[1,3]

* stephan.kambach@gmail.com, Phone: +49 345 5526263, Fax: +49 345 5527228, Institute of Biology/Geobotany and Botanical Garden, Martin-Luther-University Halle-Wittenberg, Am Kirchtor 1, 06108, Halle, Germany

1 Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108, Halle, Germany

2 Department of Community Ecology, Helmholtz-Centre for Environmental Research, Theodor-Lieser-Str. 4, 06120, Halle, Germany

3 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

4 Plant Ecology Group, Institute of Plant Sciences, University of Bern, Altenbergrain 21, CH-3013 Bern, Switzerland

5 Centre for Development and Environment, University of Bern, Hallerstrasse, 10, CH-3012 Bern, Switzerland

6 Centre for Forest Research, Université du Québec à Montréal, P.O. Box 8888, Centre-ville Station, Montréal, QC, H3C 3P8, Canada

7 Forest Ecology and Conservation Group, Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, United Kingdom

8 Geobotany, Faculty of Biology, University of Freiburg, Schaenzlestrasse 1, 79104 Freiburg, Germany

9 Institute for Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zurich, Universitaetsstrasse 16, 8092 Zurich, Switzerland

10 CSIRO Land and Water Flagship, Private Bag 5, P.O. Wembley, Western Australia 6913, Australia
11 Univ. Grenoble Alpes, Irstea, UR EMGR, 2 rue de la Papeterie BP-76, F-38402, St-Martin-d'Hères, France
12 Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre, 60325 Frankfurt, Germany
13 Department of Systematic Botany and Functional Biodiversity, Institute of Biology, University Leipzig, Leipzig, Germany
14 Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden
15 Forest Ecology and Restoration Group, Department of Life Sciences, Universidad de Alcalá, 28805 Alcalá de Henares, Spain
16 University Stefan cel Mare of Suceava, str. Universitatii nr. 13, Suceava, 720229, Romania
17 Biogeco, INRA, Univ. Bordeaux, F-33610, Cestas, France BIOGECO, INRA, University of Bordeaux, Cestas, France
18 Swedish University of Agricultural Sciences, Umeå, Sweden
19 Biogeco, INRA, Univ. Bordeaux, F-33610, Cestas, France
20 School of Biological Sciences, Royal Holloway University of London, Egham, Surrey, TW20 0EX, United Kingdom
21 Forest Research Institute Baden-Wurttemberg, Wonnhaldestrase 4, 79100 Freiburg, Germany
22 Natural Resources Institute Finland, Natural Resources and Bioproduction, (LUKE), P.O. Box 2, 00791 Helsinki, Finland
23 Department of Earth and Environmental Sciences, University of Leuven, Celestijnenlaan 200E Box 2411, 3001 Leuven, Belgium
24 Earth and Life Institute, Environmental Sciences, Université catholique de Louvain, Belgium
25 Ghent University, Forest & Nature Lab, Department of Forest and Water Management, Geraardsbergsesteenweg 267, 9090 Gontrode, Belgium

5.1 | Summary

1. For decades, ecologists have investigated the effects of tree species diversity on tree productivity at different scales and using different approaches ranging from observational to experimental study designs. Observational approaches using national forest inventories cover large environmental and compositional gradients, which leads to a high representativeness but might mask the signals of tree diversity. In contrast, tree diversity experiments have very few confounding factors, but a limited representativeness and are still very young (< 20 years). Networks of comparative plots in established forests selected to control for potentially confounding environmental conditions (forest "exploratories") can be regarded as an intermediate approach between purely observational and experimental approaches, designed to combine the strengths of both. Results across these three approaches show variable relationships between tree species diversity and productivity. To date, no study has yet investigated specifically whether the obtained relationships are consistent for the same set of species and environmental conditions.

2. We compiled data from five national forest inventories (16,773 plots), six tree species diversity experiments (584 plots) and six networks of comparative plots (169 plots), distributed across Europe. Random- and mixed-effects models were applied to compare the species-specific growth of 64 tree species in mixed versus monospecific plots between the different approaches and between the different diversity experiments and forest types.

3. We found a general positive effect of tree species mixing on species growth (16% on average). This effect was not related to the number of admixed species and nonsignificant in tree diversity experiments. Species-specific responses to mixing were inconsistent between all three research approaches, even after including only those plots that shared the same mixtures compositions and forest types.

4. *Synthesis*. Each of the tree species investigated showed either a positive or nonsignificant response to species mixing in at least one of the three approaches. Our results thus highlighted that forest managers should consider the results from the three approaches in order to select species mixtures that maximize positive forest biodiversity and functioning relationships.

Key-words: Biodiversity, Ecosystem functioning, FunDivEurope, Tree growth, National forest inventories, Productivity, Species richness, Synthesis, TreeDivNet

5.2 | Introduction

The provisioning of ecosystem services beneficial to human well-being strongly relies on plant diversity (Cardinale *et al.* 2012). Decreases in primary producer diversity can impact ecosystem functioning and decrease ecosystem productivity and stability (Cardinale *et al.* 2012; Hooper *et al.* 2012), a phenomenon especially well studied in grassland ecosystems (e.g. Tilman *et al.* 1997; Reich *et al.* 2012; Isbell *et al.* 2015). In forest ecosystems, systematic research on the effects of species mixing on wood production dates back to the foundations of modern forestry (Hartig 1791). Current global synthesis studies concluded that, across the different forest biomes, a positive relationship between tree diversity and stand productivity prevails (Zhang *et al.* 2012; Scherer-Lorenzen 2014; Liang *et al.* 2016).

The relationship between tree diversity and productivity has already been studied using different research approaches (Table 1), starting with the analysis of forest inventories (Hartig 1791; Schwappach 1912; Wiedemann 1943), followed by silvicultural trials and tree diversity experiments (Koricheva 2002; Scherer-Lorenzen *et al.* 2005; Pretzsch 2005; Bruelheide *et al.* 2014; Tobner *et al.* 2014; Verheyen *et al.* 2016) and more recently the selection of comparative plots in mature forests (Fischer *et al.* 2010; Bruelheide *et al.* 2011; Baeten *et al.* 2013). Forest inventories usually cover large numbers of uniformly distributed plots across multiple forest types and large environmental gradients. Tree diversity experiments, in contrast, consist of spatially restricted, replicated plantations of different tree species compositions and levels of tree species diversity that show minimal variation in environmental conditions. Comparative study plots (Bruelheide *et al.* 2011) or “exploratories” (Fischer *et al.* 2010) consist of survey plots within mature forests selected to comprise of replicated levels of tree species diversity and compositions while at the same time controlling for differences in community structure and environmental conditions. They can thus be regarded as an intermediate approach that combines aspects of forest inventories and tree diversity experiments.

Regardless of the approach applied, most previous research on forest diversity-productivity relationships focussed on the effects of tree species diversity on the productivity of the community (e.g. Homeier *et al.* 2010; Paquette & Messier 2011; Vilà *et al.* 2013; Ruiz-Benito *et al.* 2014; Jucker *et al.* 2016; Liang *et al.* 2016). In theory, any positive effect of species diversity could stem from either positive interactions between the co-occurring species (complementarity effects, Loreau & Hector 2001) or from the admixing of one or few exceptionally productive or dominating species (selection effects, Loreau & Hector 2001). Depending on the forest ecosystem, species-specific growth responses to increasing tree diversity can be consistently positive (Liang *et al.* 2015; Chamagne *et al.* 2017) or variable, depending on the species and context (Jucker *et al.* 2014b; Ratcliffe *et al.* 2015; Tobner *et al.* 2016; del Río *et al.* 2017). It is unclear to what extent these differences in species responses to tree diversity are caused by differences in species-specific characteristics (Fichtner *et al.* 2017; Williams *et al.* 2017) or differences in study design. Comparing species-specific responses to mixing between the different research approaches could help to determine which species generally benefit, suffer or show divergent responses to increases in tree species diversity. Restricting these comparisons to only the set of tree species and forest types that are shared between research approaches should furthermore reduce the confounding effects of species compositions and large scale environmental context-dependency and leave mainly the effects of local environmental context-dependency and differences in stand structure.

In the FunDivEUROPE research network (functional significance of forest diversity in Europe, see Baeten *et al.* 2013), all three previously described approaches (experiments, exploratories and inventories) were applied throughout the European continent to study the effects of tree diversity on forest ecosystem functioning. The three approaches partly overlap in their species pools, although there are differences in species compositions as well as successional, structural, climatic and edaphic plot conditions. Syntheses across all three approaches can thus be applied to test whether most tree species respond consistently to species mixing. Identifying tree species that display consistent responses between different approaches and different forest types would furthermore allow isolating general patterns from context-dependent effects.

With this study, we provide a first comparison of the growth response of a large set of tree species to species mixing across three distinct research approaches (tree diversity experiments, networks of comparative plots and forest inventories). We tested the following hypotheses: (H1) across all species and research approaches, tree species growth is higher in mixed than in monospecific tree communities, (H2) across all species and research approaches, the effect of tree species mixing on species growth linearly increases with the number of admixed tree species (two-, three- or higher species mixtures), (H3) the effect of tree species mixing on growth varies between species (ranging from negative for some to positive for others), but species-level responses to mixing are consistent across approaches. We furthermore hypothesized that restricting the comparison of different research approaches to only those community compositions that are shared between different approaches will increase the consistency in tree species responses to species mixing (H4). The findings of this study should deepen our understanding of the species, environmental conditions and research designs for which consistent positive diversity-ecosystem functioning relationships can be expected.

5.3 | Material and Methods

Within the framework of the European FunDivEUROPE project (www.fundiveurope.eu), the significance of forest biodiversity for ecosystem functioning across Europe was investigated with three complementary research approaches (tree diversity experiments, networks of comparative plots in established forests and forest inventories). All approaches share a similar subset of tree species and forest types and were established in regions with similar climatic conditions (see Appendix S5.1, S5.2, S5.3 and S5.4 in Supporting Information and Baeten *et al.* 2013). The approaches differed in their representativeness of existing mature forests, the comprehensiveness of the studied tree species and environmental gradients and their extent of potentially confounding effects that could mask the effects of tree species diversity ("orthogonality", see Table 5.1, Figure 5.1 and Nadrowski *et al.* 2010,).

RESEARCH APPROACHES

The experimental research approach contained growth measurements from six European tree diversity experiments, which individually covered species richness gradients from one up to six tree species, with different mixtures replicated at each level of species richness. Detailed information on the design and tree species composition of each diversity experiment is reported in Appendix S5.1 and on www.treedivnet.ugent.be. Tree sizes were measured in 2014 and reported as either tree's diameter at breast height and the derived basal area (in 114 plots of the Satakunta, 96 plots of the Kreinitz and 32 plots of the BIOTREE experiment), tree height (in 256 plots of the ORPHEE experiment) or diameter at ground height (in 42 plots of the FORBIO - Zedelgem and 44 plots of the FORBIO - Gedinne experiment).

The exploratory research approach contained a network of 209 comparative study plots that were established in six different European forest types. In each forest type, between three and five regionally common, and from a forestry perspective, important tree species were selected as target species. Plots representing species richness gradients from one up to five target tree species were established in 2011. Similar to the experimental approach, different compositions per tree species richness level were chosen to guarantee that diversity effects were not confounded with the effects of diluting individual species in plots of higher species richness and the plots were selected to minimize any co-variation between environmental conditions (e.g. geology, soil texture and depth and topography) and tree species richness and composition. The study design as well as the forest characteristics and tree species compositions are described in Appendix S5.1, S5.2, S5.3, S5.4 and in Baeten *et al.* (2013). Within each plot, all trees with a dbh of more than 7.5 cm were mapped and identified. From a subset of trees, wood core samples were taken and, based on radial stem increments between 1999 and 2010, the mean annual increase in basal area per tree was calculated (m^2 ha^{-1} $year^{-1}$, S5.7, see Jucker *et al.* 2014a). The number of plots per forest type were: beech forest (24), boreal forest (28), hemiboreal forest (25), Mediterranean coniferous forest (33), mountainous beech forest (26) and thermophilous deciduous forest (33). We calculated for each plot the proportion that was covered by each tree species and classified each plot as either a monospecific, two-, three- or higher species mixture whereby the most dominant species must cover more than 90% and none of the ‘non-dominant’ species more than 10% of a plot`s summed basal area.

The inventory research approach contained harmonised forest plots from five national forest inventories (Finland, Sweden, Germany, Belgium - Wallonia and Spain) that had been surveyed at least twice. Details can be found in S5.5 and in Ratcliffe *et al.* (2016). In short, for all trees with a dbh of 10 cm or more we extracted the tree status (ingrowth, survivor, dead due to natural mortality or harvesting) and basal area (expressed as m^2 ha^{-1}) from the two most recent survey dates. We discarded all plots with indications of harvesting activities between survey dates. Tree species names were harmonized following the Atlas Florae Europaeae (Kurtto *et al.* 2013). Within each plot, we calculated the proportion of basal area that was covered by each tree species. Analogue to the exploratory approach, we classified each plot as either a monospecific, two-, three- or higher species mixture. After discarding all plots that did not meet these criteria, we retained 47,754 plots in the inventory dataset (see Appendix S4 for a more detailed description of the classification criteria).

ENVIRONMENTAL DATA

For each plot of the three research approaches, we extracted mean annual temperature, temperature seasonality (standard deviation of mean monthly temperatures), annual precipitation and precipitation seasonality (standard deviation of mean monthly precipitation) from the WorldClim dataset (interpolated from measurements taken between 1960 to 1990 and at a spatial resolution of 1 km², Hijmans *et al.* 2005) and the slope from the GTOPO30 – digital elevation model with a spatial resolution of 1 km² (data available from the U.S. Geological Survey).

DATA PREPARATION

For each plot of the experimental, exploratory and inventory approach, we calculated for every target/dominant species the yearly summed increase in basal area, dbh, tree height or diameter at ground height (based on the respective growth measurement). These summed growth estimates were divided by the number of trees in the experiments and by the summed basal area (m^2 ha^{-1}) of the respective tree species in the exploratory and inventory approach to obtain growth estimates (hereafter "species growth") that are not biased by potentially uneven species proportions.

Within each forest type and tree diversity experiment we quantified the effect of species mixing on species growth as the mean log response ratio, defined as species growth in mixed divided by species growth in monospecific plots of comparable stand conditions (i.e. within the same dataset and forest type). In the exploratory approach, no monospecific plots of *Acer pseudoplatanus* L. were found in the beech forest and no monospecific plots of *Betula spec.* and *Quercus robur* L. were found in the hemiboreal forest. For these three species, we could not calculate the effect sizes in the respective forest types which, thus, reduced our exploratory dataset to 169 plots.

In the inventory approach, mixed and monospecific plots within the same forest type could differ considerably in stand conditions (e.g. in climate, tree community structure and edaphic conditions). To partly control for these potentially confounding differences, we first assigned pairs of monospecific and mixed plots that were most similar regarding stand and environmental conditions and subsequently calculated the effect size for each pair of plots. The dissimilarity in stand and environmental conditions was quantified as the Euclidean distance in normalized plot-level values (i.e. subtracted by the mean and divided by the standard deviation) of mean annual temperature, temperature seasonality, annual precipitation, precipitation seasonality, slope and the sum and coefficient of variation of trees' basal area (m^2 ha^{-1}). The pairs of most similar mixed and monospecific plots (i.e. with the smallest Euclidean distances) were selected via a nearest neighbour matching algorithm (Ho *et al.* 2007, 2011) that minimized, within each forest type, the summed Euclidean distances. This was done for each species separately, to compare species growth in mixed versus monospecific plots. A three-species mixture could thus be paired with up to three monospecific plots of its component species (note that a monospecific plot could only be assigned to one mixture plot). To eliminate comparisons between very different stand conditions, we discarded all plot pairs with distance values that were above the 90% percentile of all distances (see Figure S5.9). The locations of the remaining 16,773 plots are shown in Figure S4.6. All plots were assigned to one of the following forest types, listed in the EEA Technical Report 9 (Barbati *et al.* 2007): acidophilous oak and oak-birch forest (104 plots), alpine coniferous forest (615), beech forest (475), boreal forest (2,440), broadleaved evergreen forest (2,129), floodplain forest (20), hemiboreal forest and nemoral coniferous and mixed broadleaved-coniferous forest (1,391), plantations and exotic forest (1,088), Mediterranean coniferous forest (6,098), mesophytic deciduous forest (582), mountainous beech forest (426), non-riverine alder, birch or aspen forest (254), mire and swamp forest (204) or thermophilous deciduous forest (947). Because the survey dates and the methods applied to measure tree growth differed between the different national forest inventories, we noted the country of each mixed and monospecific plot to later statistically account for it.

In order to narrow down the comparisons of mixing effects to only those tree species and community compositions that were shared between the three approaches, we created three data subsets that included only those species and mixtures that were present in two datasets, i.e. (i) the experimental and exploratory, (ii) the experimental and inventory and (iii) the exploratory and inventory approach (Table S5.4).

STATISTICAL ANALYSIS

Separately for each tree diversity experiment and each forest type within the exploratory or the inventory dataset, we calculated for every tree species the separate mean log response ratio (hereafter “effect size”) of the species’ growth in either all 2-, 3- or higher species mixtures divided by the growth in the respective monospecific plots of that forest type/diversity experiment. The whole data preparation procedure up to the point of the calculation of effect sizes is briefly summarized in Appendix S5.8.

We tested hypothesis H1 (i.e. a general positive effect of tree species mixing on species growth) by testing for significance of the grand mean effect size (i.e. the intercept) with a linear random-effects model. The model included effect sizes as the dependent variable and the identity of the experiment/forest type and, in the case of the inventory approach, the countries of the compared plots, as random effects. Individual effect sizes were weighted by the inverse of the number of effect sizes for the same species in the same experiment/forest type (this number could vary when plots from species had different numbers of richness levels or different forest inventories were assigned to the same forest type). The resulting grand mean effect size was deemed significant, if the approximated 95% confidence interval (intercept ± 1.96 x standard error) did not include zero. We tested the differences between approaches by including the research approach as a categorical predictor variable in the mixed-effects model.

Hypothesis H2 (i.e. a positive effect of species richness on the species’ mean log response ratios) was tested with linear mixed-effects models that included the effect sizes as the dependent variable, species richness as the predictor variable and the identity of the forest type or experiment and, in case of the inventory approach, the countries of the compared plots as a nested random effect. In contrast to the model applied to test H1, we assigned equal weights to all effect sizes. Only in the inventory approach we weighted effects sizes by the inverse of the number of effect sizes for the same species in the same forest type (this number could vary when plots from different forest inventories were assigned to the same forest type). H2 was then tested by comparing the variance explained with the full model versus the variance explained with solely the random-effects (analysis of variance).

In order to test hypothesis H3 (i.e. the consistency in species-specific responses to mixing across the research approaches), we fitted separate mixed-effects models per approach (for the experimental, exploratory and inventory approach, respectively). These models included the identity of the tree species as a predictor variable and the random-effects structure was adapted from the model that was applied to test H1. The intercept of each model was set to zero. From each model, we then extracted the coefficient estimates for the respective tree species included. The consistency in species responses was then assessed by testing the significance of the rank-based correlation coefficients (Kendall's tau) between the coefficient estimates of species that were shared between different approaches (separately for the experiments-exploratories, experiments-inventories and exploratories-inventories comparison).

Hypothesis H4 (i.e. the proposed increase in the consistency of species responses to mixing when the comparisons of approaches were restricted to only those community compositions and forest types that are shared between the approaches) was tested analogue to H3, but this time based on data sets restricted to tree species occurring in the same compositions and forest types in the compared research approaches (listed in Table S5.4). The obtained Kendall's tau values were then compared to the tau values that were obtained from the unrestricted data sets.

All analyses were conducted in *R* (R Core Team 2013) using the following packages: *ggplot2* for graphical representations (Wickham 2009), *cluster* for distance matrix calculations (Maechler *et al.* 2015), *data.table*, *dplyr* and *reshape2* for data wrangling (Wickham 2007; Dowle *et al.* 2015; Wickham & Francois 2015), *MatchIt* for finding pairs of similar mixed and monospecific plots (Ho *et al.* 2011)*, lme4* for calculating linear random- and mixed-effects models (Bates *et al.* 2015) and *raster* for extracting the *WorldClim* data (Hijmans 2013).

5.4 | Results

(H1) When calculated across all three research approaches (experiments, exploratories and inventories), the grand mean effect size of species mixing (i.e. the average log response ratio of species growth in mixed compared to monospecific plots) was significantly positive (approximated 95% confidence interval: 0.05 to 0.25). On average, species showed 16% higher growth in mixed compared to monospecific plots. When calculated separately for each research approach, both the inventory and exploratory dataset yielded significantly positive mean effect sizes (on average, species growth was 27% and 20% higher in mixed compared to monospecific plots of the exploratory and inventory approach, respectively, Figure 5.2) whereas the mean effect size of the experimental approach was non-significant (on average, species growth was 1% higher in mixed compared to monospecific plots, Figure 5.2). In the experimental approach, none of the mean effect sizes (average species log response ratios) of the individual diversity experiments was significantly different from zero. In the exploratory approach, significantly positive mean effect sizes were found in Mediterranean coniferous, thermophilous deciduous and boreal forests. In the inventory approach, significantly positive mean effect sizes were found in beech, thermophilous deciduous, alpine, Mediterranean coniferous, boreal and mountain beech forests.

(H2) Including tree species richness as a predictor variable did not explain a significant amount of variation in species' effect sizes ($F_{df: 1, 299.75} = 1.6$, $p = 0.21$).

(H3) Between the different research approaches, tree species responses to mixing (i.e. the model coefficient estimates) were highly inconsistent (Figure 5.3). All Kendall's tau values ranked between 0.55 and 0.94 and were nonsignificant (p-values ranged from 0.55 to 0.94). *Fraxinus excelsior* L. was the only species to exhibit consistent, and positive, effects sizes in all three research approaches (Figure 5.3)

(H4) Restricting the comparisons to only those species compositions and forest types that were shared between the compared research approaches did not lead to stronger correlations between species' coefficient estimates of different approaches (Figure S5.10). Kendall's tau values ranged from -0.2 to -0.06 and the respective p-values ranged from 0.72 to 0.84.

5.5 | Discussion

In this study, we compiled tree growth data from three European research initiatives that followed different research approaches (tree diversity experiments, networks of comparative "exploratory" plots in established forests and national forest inventories) to summarize the effects of tree species mixing on the growth of 64 tree species.

Based on this extensive data set, we conducted, to our knowledge, the first study on the transferability of the response of tree species growth to mixing from experiments to forest exploratories and national forest inventories. Our results confirmed our hypothesis of a general positive effect of tree species mixing on species growth across the three research approaches, although this effect was nonsignificant in the experiments. This finding is in accordance with the meta-analysis of (Piotto 2008) who also found that tree species generally exhibit higher growth in mixed compared to monospecific communities. In the exploratory and inventory data set, tree species showed, on average, an increase of 27% and 20% in growth in mixed as compared to monospecific stands. Studies that investigated the effect of species mixing on the productivity of the whole tree community (as opposed to the growth of the individual species) reported positive effects of comparable magnitude. Tree communities exhibited a 21% higher productivity in mixed as compared to the respective monospecific stand in the Spanish forest inventory (Ruiz-Benito *et al.* 2014) and 24% higher productivity across the national forest inventories of France, the Netherlands, Spain, Sweden and Switzerland (Vilà *et al.* 2013).

Previous analyses of the published literature (Zhang *et al.* 2012), the Spanish national forest inventory (Ruiz-Benito *et al.* 2014) and a global forest dataset (Liang *et al.* 2016) all found that the productivity of the whole tree community increases with the number of mixed tree species. In our analyses of individual species, however, we could not find such an increase in the magnitude of the mixing effect with the number of admixed tree species.

Regarding the exploratory approach, our results confirmed the findings of Jucker *et al.* (2014a), who previously analysed the same exploratory dataset, and also found positive effects of species mixing on plot productivity in the Mediterranean coniferous, thermophilous deciduous and boreal forests type. Our findings are also in line with studies that investigated the same inventory dataset and found positive effect of tree diversity on the productivity of the whole tree community (Ratcliffe *et al.* 2016; Madrigal-González *et al.* 2016; Ruiz-Benito *et al.* 2017), although we investigated the effects on individual species and manipulated the inventory dataset to make it compatible to the exploratories.

Our results further suggested that species mixing mostly benefitted those species that grew in forest types in a relatively cold climate (boreal and alpine forests) or a relatively hot climate (Mediterranean coniferous and thermophilous deciduous forests). These observations are in line with an analysis of an eastern Canadian forest inventory dataset that likewise found stronger positive effects of tree diversity on stand productivity in boreal as compared in temperate forests (Paquette & Messier 2011). Together, these findings broadly support the stress-gradient hypothesis, stating that positive interactions prevail in more stressful conditions (e.g. cold or dry), resulting in higher relative diversity effects than in more benign conditions (Forrester & Bauhus 2016). We found consistent species responses to mixing between the exploratory and inventory approach only for those three forest types of the most stressful climatic conditions. However, for the remaining three forest types that were shared between both approaches and found in intermediate conditions, we found no consistency in the significance or even direction of the mixing effect. This limited transferability of mixing effects between approaches, already indicated that scaling of diversity effects across approaches might problematic.

Consequently, we found that species-specific responses to mixing were largely inconsistent between all three approaches, even after restricting the data sets to plots of only those species compositions and forest types that were shared between the different approaches. These observed inconsistencies likely resulted from unaccounted but influential factors that shaped forest diversity and functioning relationships, for instances any general differences in tree density, size heterogeneity and the successional status (Lasky *et al.* 2014) between the compared research approaches. In accordance with the recent global meta-analysis conducted by Duffy *et al.* (2017), our analyses clearly showed stronger biodiversity effects in established forests than in experimental ones. Our findings thus support the conclusions of Duffy *et al.* (2017) that biodiversity effects on biomass are generally stronger in nature than in biodiversity experiments. It has previously been argued that biodiversity effects on ecosystem functioning should be much stronger in experiments than in naturally assembled ecosystems because species decline in experiments is mostly simulated to be at random, whereas in real communities, species loss is often biased towards the rare and least productive species, resulting in a weak connection between species loss and productivity (Lepš 2004). However, our results suggest the contrary, i.e. stronger relationships in naturally assembled communities, suggesting that biodiversity-ecosystem functioning mechanisms are at least equally import in these systems as in diversity experiments (Duffy *et al.* 2017). interactions among coexisting species can be strong enough to produce positive diversity-growth relationships even under natural assembly whereas diversity experiments could rather underestimate the effects of tree diversity on tree growth.

Differences in the prevailing climatic conditions can also lead to differences in observed forest biodiversity-productivity relationships (Paquette & Messier 2011; Jucker *et al.* 2016; Ratcliffe *et al.* 2017) and although the three compared approaches were established in overlapping climatic conditions they still display varied in climatic and probably also edaphic conditions.

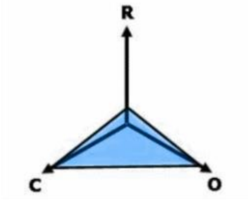
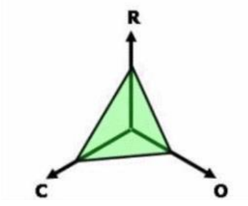
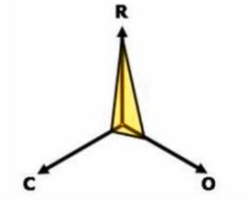
Madrigal-González *et al.* (2016) furthermore demonstrated the impact of the diversity of neighbour trees on tree growth can be mediated by an interaction effect of tree sizes and climatic conditions. More specifically, across the national forest inventories of Finland, Germany, Spain, Sweden and Belgium-Wallonia, Madrigal-González *et al.* (2016) found that smaller trees benefitted from a complementary (i.e. functionally divergent) neighbourhood only in the coldest and intermediate regions whereas larger trees benefitted from complementarity only in the warmest regions. With the approach applied in this study (i.e. the comparison of mean species growth between mixed and monospecific plots), we could not account for the potentially confounding differences in tree sizes and especially the interaction with prevailing climatic conditions.

Herbivore pressure is another factor that likely varied between the three approaches. Except for the Satakunta site, all tree diversity experiments were fenced to exclude game species and safeguard the successful establishment of all planted trees. In the inventory, and even more in the exploratory approach, the juvenile trees are exposed to pressure by game herbivores which is known to be affected by tree species richness (Milligan & Koricheva 2013; Ohse *et al.* 2017). The effects of tree diversity on forest functioning are scale-dependent, meaning that significance can change with the size of the surveyed forest plots (Wang *et al.* 2016). Inconsistencies in species-specific responses could thus partly result from differences plot size and spatial extent between the compared research approaches.

In summary, all of the proposed factors might have contributed to the inconsistency of species-specific responses to mixing between tree diversity experiments and established forests. On the one hand, these results impede clear recommendations for forest owners on how to jointly maximize forest diversity and productivity. On the other hand, our results unequivocally demonstrated that not even one of the 64 investigated tree species generally suffers from species mixing. Beside the hemiboreal forests in the inventory approach, most tree species were, on average, either not significantly or even positively affected by species mixing. We thus concluded that many, if not most, monospecific stands can be diversified without negative or with positive effects on wood production.

Future research will be needed to answer (1) what are underlying causes that lead to different diversity-functioning relationships between observational and experimental research approaches and (2) what are the species-specific abiotic and biotic requirements that maximize the productivity in mixed and monospecific communities. These findings will be essential to devise forest management practices that can maximize synergies between wood production and the safeguarding of forest diversity in Europe (Chamagne *et al.* 2017).

Table 5.1 | Summary on the advantages, disadvantages and exemplary findings on the relationship between tree species diversity and tree growth or stand-level biomass production in three different research approaches. Figures depict the characteristics of the research approaches: **R**epresentativeness (i.e. the anticipated transferability of the findings to existing forests), **C**omprehensiveness (i.e. the number of ecosystem functions and properties that can be feasibly quantified) and **O**rthogonality (i.e. the ability to quantify the effect of tree diversity against a background of variation); Figures are based on Nadrowski *et al.* (2010) and Jucker *et al.* (2016) and published on project.fundiveurope.eu.

Research approach	Advantages	Disadvantages	Reported effects of tree diversity on productivity
Tree diversity experiments 	Solid statistical design Can include species mixtures that do not occur naturally Minimal variation in environmental characteristics Diversity orthogonal to other drivers of function Causal inference possible	Fixed number of tree species and combinations Cover only limited environmental gradients	**Global network of tree diversity experiments**: Verheyen *et al.* 2016, www.treedivnet.ugent.be **Positive:** Pretzsch 2005; Erskine *et al.* 2006; Potvin & Gotelli 2008; Haase *et al.* 2015; Fichtner *et al*. 2017 **Nonsignificant:** Nguyen *et al.* 2012; Guo & Ren 2014; Tobner *et al.* 2016 **Negative:** Firn *et al.* 2007
Comparative forest plots (exploratories) 	Controlled species composition Intermediate variation in stand characteristics Diversity as orthogonal as possible to other drivers of function Intermediate gradient in environmental conditions Can be established in mature forests	Limited number of tree species Causal inference is difficult	**Positive:** Baruffol *et al.* 2013; Jucker *et al.* 2014a **Negative:** Jacob *et al.* 2010
Forest inventories 	Large number of plots Vast geographic extend Large gradients in - Species compositions - Stand characteristics - Environmental conditions Highly representative	Large heterogeneity can confound diversity signals Design originally not developed to study biodiversity-ecosystem function relationships Causal inference not possible	**Positive:** Vilà *et al.* 2007; Paquette & Messier 2011; Vilà *et al.* 2013; Guo & Ren 2014; Ruiz-Benito *et al.* 2014; Liang *et al.* 2016; Madrigal-González *et al.* 2016; Ratcliffe *et al.* 2017; Ruiz-Benito *et al.* 2017 **Nonsignificant:** Szwagrzyk & Gazda 2007; Moser & Hansen 2009; Long & Shaw 2010; Vayreda *et al.* 2012 **Hump-shaped:** Gamfeldt *et al.* 2013 **Negative:** Mina *et al.* 2017

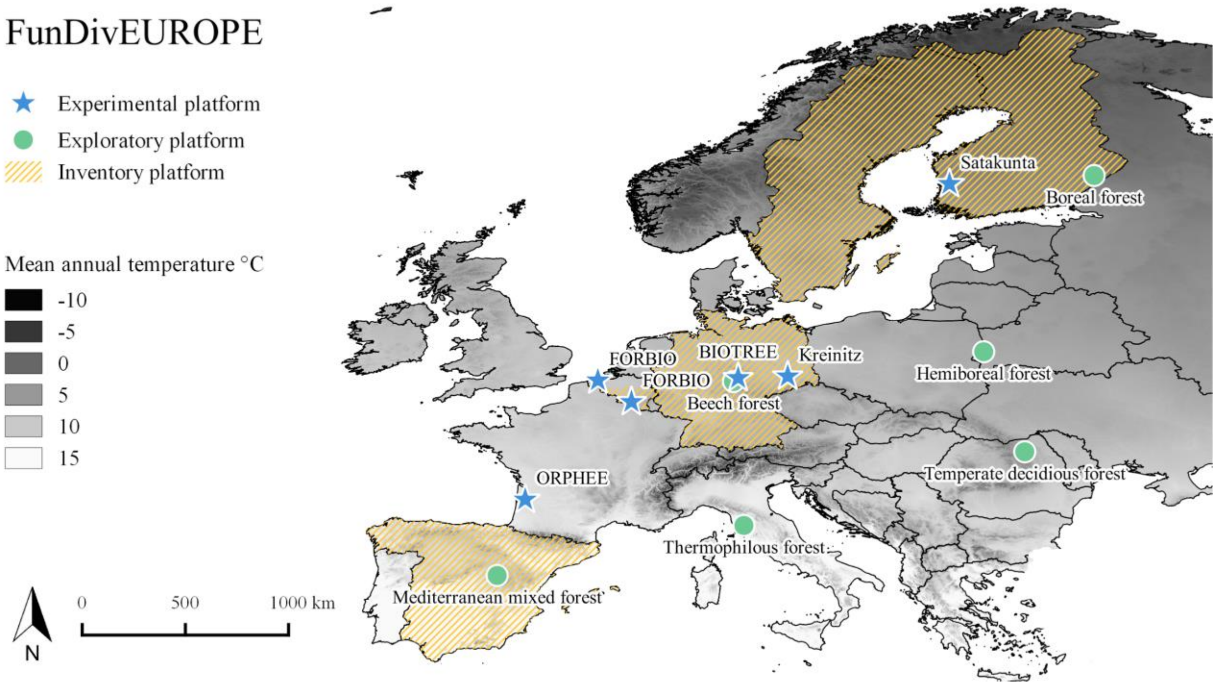


Figure 5.1 | Location of the research approaches compiled in this study. Shaded countries: national forest inventories (16,773 plots), stars: tree diversity experiments (584 plots), black dots: forest exploratories (169 plots).

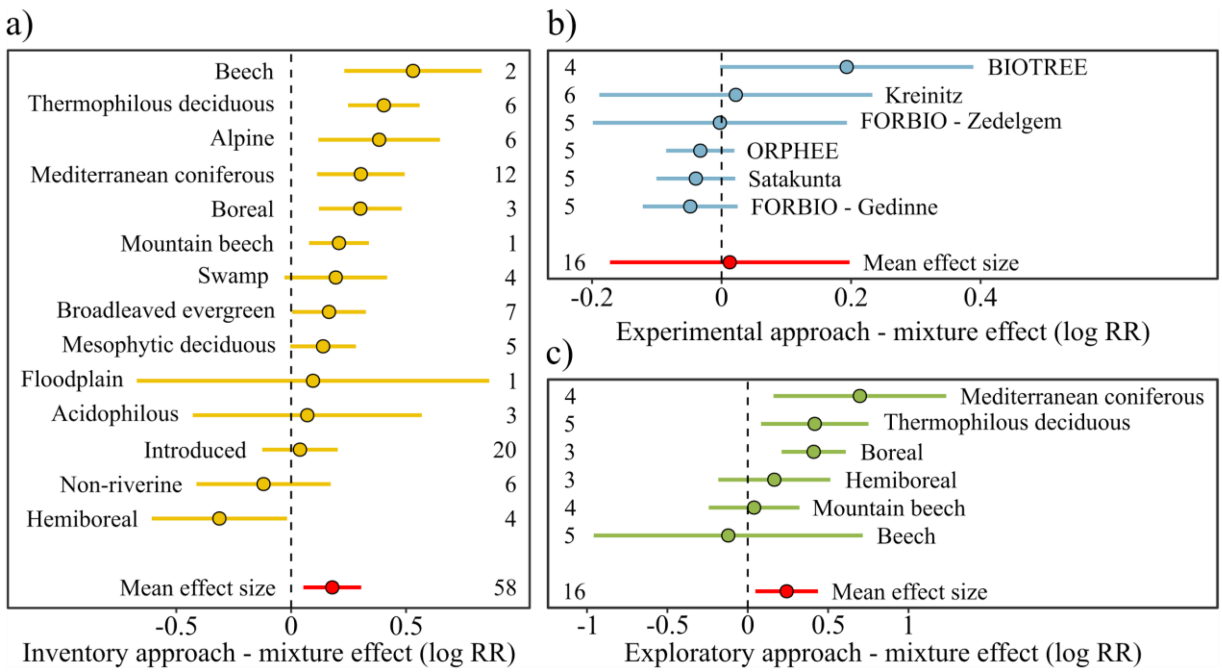


Figure 5.2 | Mean effect sizes (log response ratios) of tree species growth in mixed compared to monospecific plots averaged per forest type/tree diversity experiment in the three different research approaches: a) forest inventories, b) tree diversity experiments and c) forest exploratories. Numbers denote the number of tree species for which effect sizes could be calculated. Different forest types/diversity experiment could overlap in the analysed tree species. Thus, the species of the grand mean effect sizes are lower than the summed species numbers.

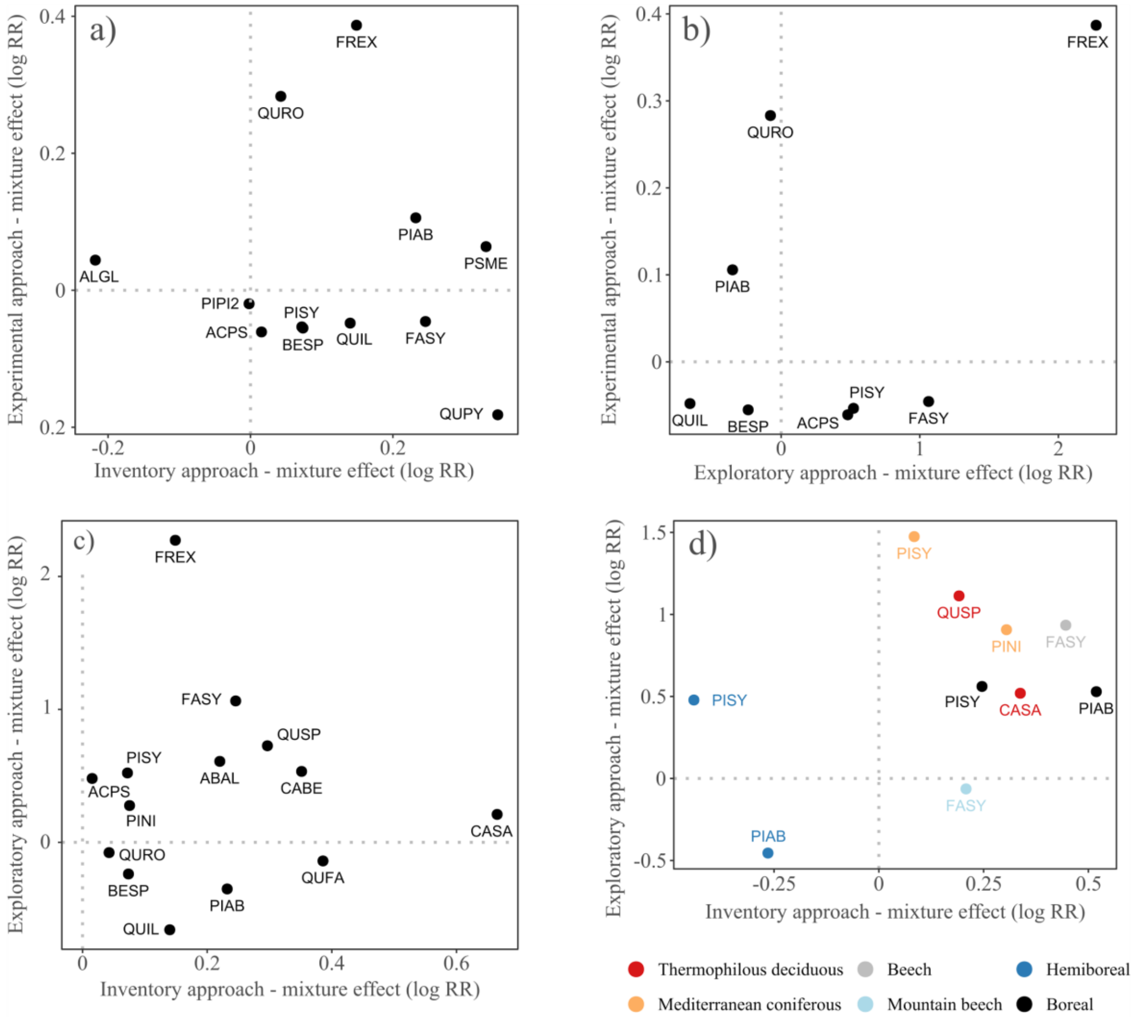


Figure 5.3 | Comparison of tree species mean effect sizes (log response ratios) of growth in mixed compared to monospecific plots obtained from three different research approaches (experimental, exploratory and inventory approach). Depicted are the mean effect sizes of only those species that were shared between the compared research approaches (a: experiments versus inventories, b: experiments versus exploratories, c: exploratories versus inventories, d: exploratories versus inventories when species responses were separated by forest type). Abbreviations: ABAL: *Abies alba* Mill., ACPS: *Acer pseudoplatanus* L., BESP: *Betula spec*., ALGL: *Alnus glutinosa* (L.) Gaertn., CABE: *Carpinus betulus* L., CASA: *Castanea sativa* Mill., FASY: *Fagus sylvatica* L., FREX: *Fraxinus excelsior* L., PIAB: *Picea abies* (L.) H. Karst., PINI: *Pinus nigra* J. F. Arnold, PIPI2: *Pinus pinea* L., PISY: *Pinus sylvestris* L., PSME: *Pseudotsuga menziesii* (Mirb.) Franco, QUFA: *Quercus faginea* Lam., QUIL: *Quercus ilex* L., QUPY: *Quercus pyrenaica* Willd., QURO: *Quercus robur* L., QUSP: *Quercus spec* - combines *Q. petraea* and *Q. pubescens Willd. (Q. Humilis)* (Table S4.2).

5.6 | Author Contributions, Acknowledgements and Data Accessibility

M.S.-L., K.V., J.K., H.J., H.B., H.A., D.A.C., B.M., Q.P. designed the FunDivEUROPE experimental and exploratory platforms. S.R., P.R.-B., A.L., G.K., J.D., M.A.Z. contributed to the harmonisation of the NFI data sets. The original ideas of this study were conceived by E.A., L.G., M.S.-L., F.v.d.P., H.B., O.B., S.B.-G., J.H., T.J., G.K., S.K., S.M., C.N., A.P., Q.P., F.R., S.R., and P. R.-B. during an sDiv workshop that was organized by E.A. and L.G. The data was collated by O.B., D.A.C., J.D., T.J., J.H., J.K., Q.P. and T.V.d.P. S.K. performed the statistical analyses with the support of H.B., J.H., G.K., F.v.d.P., S.R. and P.R.-B. S.K. wrote the first draft of the manuscript to which all co-authors contributed critically and gave the final approval for publication.

This paper is a joint effort of the working group sFundivEurope kindly supported by sDiv, the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118). The FunDivEUROPE project received funding from the European Union's Seventh Programme (FP7/2007–2013) under grant agreement No. 26517.

The work of S. Kambach was funded by the Helmholtz Centre for Environmental Research – UFZ, the Helmholtz Research School for Ecosystem Services under Changing Land-use and Climate (ESCALATE), the Martin-Luther-University Halle-Wittenberg and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. We thank the MAGRAMA for access to the Spanish Forest Inventory, the Johann Heinrich von Thünen-Institut for access to the German National Forest Inventories, the Finnish Forest Research Institute (METLA) for making the Finnish NFI data available, the Swedish University of Agricultural Sciences for making the Swedish NFI data available, and Hugues Lecomte, from the Walloon Forest Inventory, for access to the Walloon NFI data. We thank all the people and organizations that dedicated their time and work to plan and maintain the tree diversity experiments, the exploratory forest plots and the network of tree diversity experiment (TreeDivNet, www.treedivnet.ugent.be/index.html).

Information on the availability of the National Forest Inventory datasets can be found on the following websites: Finland - www.metla.fi/ohjelma/vmi/info-en.htm, Germany - bwi.info/?lang=en, Spain - ww.magrama.gob.es/es/desarrollo-rural/temas/politica-forestal/inventario-cartografia/ inventario-forestal-nacional, Sweden -www.slu.se/nfi - iprfw.spw.wallonie.

6 | Study 2 - The Impact of Tree Diversity on Different Aspects of Insect Herbivory Along a Global Temperature Gradient - a Meta-Analysis

Stephan Kambach[1, 2], Ingolf Kühn[1, 2, 3], Bastien Castagneyrol[4], Helge Bruelheide[2, 3]

1 Department of Community Ecology, Helmholtz-Centre for Environmental Research, Halle (Saale), Germany

2 Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

3 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

4 BIOGECO, INRA, Univ. Bordeaux, 33610 Cestas, France

6.1 | Abstract

Forests with higher tree diversity are often assumed to be more resistant to insect herbivores but whether this effect depends on climatic conditions is so far poorly understood. In particular, a forest's resistance to herbivory may depend on mean annual temperature (MAT) as a key driver of plant and insect phenology. We carried out a global meta-analysis on regression coefficients between tree diversity and four aspects of insect herbivory, namely herbivore damage, abundance, incidence rate and species richness. To test for a potential shift of tree diversity effects along a global gradient of MAT we applied mixed-effects models and estimated grand mean effect sizes and the influence of MAT, experimental vs. observational studies and herbivores diet breadth. There was no overall effect of tree diversity on the pooled effect sizes of insect herbivore damage, abundance and incidence rate. However, when analysed separately, we found positive grand mean effect sizes for herbivore abundance and species richness. For herbivore damage and incidence rate we found a significant but opposing shift along a gradient of MAT indicating that with increasing MAT diversity effects on herbivore damage tend towards associational resistance whereas diversity effects on incidence rates tend towards associational susceptibility. Our results contradict previous meta-analyses reporting overall associational resistance to insect herbivores in mixed forests. Instead, we report that tree diversity effects on insect herbivores can follow a biogeographic pattern calling for further in-depth studies in this field.

6.2 | Introduction

Insect herbivores can compromise the functioning of forest ecosystems (Boyd *et al.* 2013). Insect herbivory is controlled by top-down mechanisms involving natural enemies of herbivores (Price *et al.* 1980), bottom-up mechanisms including tree defences (Coley 1983) and associational effects provided by tree diversity (Schuldt *et al.* 2010, Schuldt *et al.* 2015). So-called associational effects occur if herbivory on individual trees is influenced by the identity and density of neighbouring trees (Underwood *et al.* 2014) and are assumed to be key regulators of herbivory (Barbosa *et al.* 2009, Underwood *et al.* 2014). An increase in tree diversity is often reported to decrease herbivore pressure (associational resistance - AR; Jactel & Brockerhoff 2007, Barbosa *et al.* 2009, Castagneyrol *et al.* 2014), but examples for an opposite relationship can also be found (associational susceptibility - AS; White & Whitham 2000, Barbosa *et al.* 2009).

Associational resistance, on the one hand, occurs if a higher diversity of tree species reduces the availability of host trees for specialist herbivores because of reduced resource availability and reduced encounter rates of herbivores and hosts (resource dilution effects; Otway *et al.* 2005, Hambäck *et al.* 2014). Mixtures of host and non-host tree species can also reduce the success of herbivores to detect suitable resources, both visually (Castagneyrol *et al.* 2013) and olfactorily (Jactel *et al.* 2011).

Associational susceptibility, on the other hand, occurs if a population of herbivores builds up on a more preferred host species and then ‘spills over’ to admixed less palatable species (Brown & Ewel 1987, White & Whitham 2008). Generalist herbivores might furthermore benefit from a more diversified nutrition (Raubenheimer & Simpson 1999, Unsicker *et al.* 2008) as suggested by the dietary mixing hypothesis (Tahvanainen & Root 1972, Bernays et al. 1994).

When accounting for the whole community of herbivores that are able to attack a particular tree, any change in the proportion of generalist vs. specialist herbivores may then shift the balance between AS and AR. Yet, the proportion of specialized insect herbivores was reported to increases towards lower latitudes (Forister *et al.* 2015). As such, it is likely that the direction of associational effects changes along latitudinal gradients.

Several other key features of plant-herbivore interactions are known to vary with latitude. Numerous studies also found that global gradients exist for plant species richness (Kreft & Jetz 2007, Gillman *et al.* 2015), herbivore species richness (Price *et al.* 1998, Hillebrand 2004, Salazar

& Marquis 2012), plant defences (Moles *et al.* 2011, Pearse & Hipp 2012), herbivore pressure (Adams & Zhang 2009, Pennings *et al.* 2009, Salazar & Marquis 2012), leaf herbivory (Kozlov *et al.* 2015) and trophic interactions (Rodríguez-Castañeda 2013). Since temperature is a key driver of herbivore development and abundance (Bale *et al.* 2002), consumption rates (Lemoine *et al.* 2014) and host-plant choices (Clissold *et al.* 2013) it is likely to influence these plant-herbivore interactions (Clissold *et al.* 2013, Lemoine *et al.* 2014). Yet, and surprisingly, it still remains to be tested whether the direction and strength of associational effects also change along a global gradient of mean annual temperature (MAT).

In addition to the overlooked effect of MAT on associational effects, the current understanding of AR and AS suffers from several methodological and conceptual biases, including a lack of considering diversity gradients, and confusion between functional (i.e. consumption or proportion of attacked tissue) and quantitative (i.e. abundance and species richness) responses of herbivores.

Previous meta-analyses on associational effects in tree stands mainly compared single species (i.e. monocultures) with mixed stands, irrespective of species richness and species evenness (Jactel & Brockerhoff 2007, Castagneyrol *et al.* 2014). Yet, the relative share of tree species within tree stands may critically change the way herbivores perceive stand quality (Castagneyrol *et al.* 2014, Verschut *et al.* 2016). To our knowledge, no study so far has investigated whether the strength and direction of associational effects depend on the metric of tree diversity applied. We therefore compiled studies that reported on either the Shannon or Simpson diversity of tree stands reflecting the relative contribution of host concentration and relative frequency (see Underwood *et al.* 2014 and Kim & Underwood 2015 for a discussion about host concentration vs. frequency). If species proportions were not available we applied a gradient of tree species richness or, if species richness was not reported, conducted a comparison between single species and mixed tree stands.

Beside the simplification to monoculture-mixture comparisons, the meta-analyses mentioned above (Jactel & Brockerhoff 2007, Castagneyrol *et al.* 2014) also simplified the response of herbivores by pooling studies that reported on the actual damage inflicted and studies that reported on the abundance of insect herbivores, assuming that more herbivores always cause more damage. This assumption mainly holds for some herbivores such as bark-beetles, leaf miners or galls, but is more controversial for most of defoliators. For instance, herbivore abundance and the associated damage are not necessarily correlated (Rhainds & English-Loeb 2003, Stoepler & Rehill 2012) and these two aspects have been reported to sometimes respond differently to plant diversity (Barbosa *et al.* 2009). Beside, how the species richness of insect herbivores depends on tree diversity has not been summarized in previous meta-analyses. Thus, we separately analysed the response of insect herbivores to increasing tree diversity into the response of herbivore damage, abundance, incidence rate (i.e. the proportion of attacked plant tissue) and the species richness of herbivores.

In the present study, we went beyond previous meta-analyses by testing whether the direction and strength of associational effects on forest trees can partly be attributed to the MAT at the study location. Moreover, we asked whether the focus on different aspects of herbivory (i.e. the amount of insect herbivore damage, abundance, incidence rate and species richness) as well as on different metrics of tree diversity (i.e. the Shannon/Simpson diversity, the species richness and the comparison of single species vs. mixed tree stands) can lead to diverging associational effects.

Specifically, we hypothesized that

i) insect herbivory (pooled over herbivore damage, abundance and incidence rate) is negatively related to tree species diversity;
ii) insect herbivore damage, abundance, incidence rate and species richness differ in their relationship to tree species diversity;
iii) the relationships between the four aspects of insect herbivory (i.e. herbivore damage, abundance, incidence rate and species richness) and tree diversity change with mean annual temperature.

6.3 | Material and Methods

Search strategy and inclusion criteria:

We performed a scoping search with combinations of relevant search terms including the following key words: forest, tree, diversity, richness, herbivores, pest, damage, monoculture, mixture and plantation. The following literature databases were queried in August 2016: Thomson Reuters Web of Science, Google Scholar and Cab Direct. These search queries were partly conducted with the help of the program *Publish or Perish 4* (Harzing 2014). The PRISMA checklist of items to include when reporting a systematic review is shown in Appendix S6.1 (Moher *et al.* 2010) and the search protocol used for each database is fully described in Appendix S6.2. After additionally reviewing the literature cited in relevant articles, our search initially yielded 3,707 articles. The search and subsequent selection process is depicted in the flow chart Figure 6.1.

After the initial screening, we retained 1,234 articles for further in-depth examination. To be included in the present meta-analysis, an article had to report on i) either the Shannon diversity, Simpson diversity, species richness or a monoculture-mixture comparison of tree stands and ii) either the amount of herbivory sustained by trees, the abundance or incidence rate (e.g. the proportion of individuals or parts that were attacked) or species richness of insect herbivores. All causes leading to dismissing articles are fully described in Appendix S6.2.

Effect sizes

Due to the heterogeneity of studies, the compiled measurements of herbivory vastly differed in dimensions and order of magnitude. In the case of herbivore damage, studies reported the percentage of leaf area removed or percentage of crown volume damaged. In the case of herbivore abundance studies reported on the number or density of individuals, galls and egg clusters. Incidences rate were determined as percentage of trees, leaves, or branches attacked or with herbivores present. Species richness was measured as the number of insect herbivore species captured on trees or in traps.

Our main objective was to test the correlation between an increase in tree diversity and the four aspects of tree herbivory (i.e. damage, abundance, incidence rate and species richness). Since Pearson's correlation coefficient r becomes skewed as it approaches $\pm$ 1 we transformed r to Fisher's z-scores which range from $-\infty$ to $+\infty$, have the same sign as r values and are commonly applied in meta-analyses (Fisher 1921, Borenstein *et al.* 2011, Koricheva *et al.* 2013). Uncertainty for each effect size was estimated by calculating the corresponding variance estimate ($v = 1/(n\text{-}3)$, where n is the sample size (Fisher 1921).

When an article reported multiple measurements (e.g. for different taxa or different aspects of herbivory) we retained them as separate study cases within the same forest and accounted for this non-independence in the statistical analyses (see below).

When studies did not report any *r* value but provided tables or figures with information on tree diversity and any aspect of herbivory we extracted the raw data using the software *ImageJ* (Rasband 2014) and re-calculated the corresponding *r* values. This could only be achieved when there were at least four records for both tree diversity and insect response.

We extracted the mean annual temperature (MAT) for the approximate study locations from the *WorldClim*-global climate data (Hijmans *et al.* 2008). In order to test for potentially confounding factors, we furthermore distinguished between experimental (i.e. plantations with purposefully manipulated tree diversity) and observational studies (in either semi-natural forests or forests that had not been planted to test for any diversity effects) and classified the diet breadth of herbivores as being either specialists that feed on a single species or genus of tree species or generalists which can utilize a wide range of tree species.

Statistical Analyses

We first tested the grand mean effect size corresponding to the overall effect of increasing tree diversity on the pooled effect sizes of herbivore damage, abundance and incidence rate. Using a random-effect meta-analysis model (Koricheva *et al.* 2013) with restricted maximum likelihood we calculated the model intercept, (i.e. averaged Fisher's *z*-scores) and the corresponding bootstrap confidence interval (CI). Effect sizes were weighted based on the inverse of their variance estimate. Since multiple effect sizes from the same forest location cannot be considered as fully independent we incorporated a hierarchical random-effect structure with the single effect sizes nested within forest locations. The grand mean effect size was considered statistically significant if the 95% CIs did not include zero. We also estimated the amount of residual heterogeneity (τ^2; Koricheva *et al.* 2013) and tested whether effect sizes displayed significant between-study heterogeneity by applying the weighted Cochran's Q-test. To test whether effect sizes depended on the aspect of herbivory we pooled all effect sizes on herbivore damage, abundance, incidence rate and species richness of herbivores, included this aspect of herbivory as a moderator in the previous random-effect model and calculated the test statistic for the omnibus test of model coefficients (Q_M). We then split the dataset based on the aspect of herbivory, applied the intercept only-model from above and calculated for each aspect separately the grand mean effect size, τ^2 and Q-test statistic.

For each aspect of herbivory, we then tested the correlation between effect sizes and MAT, the type of study (experimental vs. observational) and the diet breadth of insect herbivores.

We thereby included these variables as moderators in the previous random-effect models and tested the significance of the obtained parameter estimates against a normal distribution. To exclude statistically insignificant moderators and to yield robust, most parsimonious models, we performed a backward selection based on error probabilities ($\alpha = 0.05$) calculated with maximum likelihood. The parameter estimates of the resulting four minimal adequate models were then calculated with restricted maximum likelihood. For each model we separately calculated τ^2, Q_M and the test statistic for the amount of residual (i.e. unexplained) heterogeneity (Q_E, weighted Cochran's Q-test). The amount of variation in effect sizes that predictors accounted for was finally calculated as $Q_M / (Q_M + Q_E)$.

To check for publication bias resulting from the omission of extreme or unlikely results, we visually inspected funnel plots and tested for their asymmetry by applying Egger's regression test (Sterne & Egger 2005, Borenstein *et al.* 2011). We furthermore tested whether the year of publication correlated with effect sizes or MAT at the study location by calculating Pearson correlation coefficients and the corresponding test statistics. We finally checked whether effect sizes correlated with the reported metric of tree diversity (i.e. Shannon/Simpson diversity, species richness and single vs. mixed species stands) by including this moderator in the previous random-effect model and calculating Q_M.

All statistical analyses were conducted in *R* (R Core Team 2013) using the packages *metafor* for meta-analyses (Viechtbauer 2010), *vegan* for calculating the Shannon diversity (Oksanen *et al.* 2011), *raster* for extracting the *WorldClim* data (Hijmans 2013) and *ggplot2* for graphical representations (Wickham 2009).

6.4 | Results

The final dataset consisted of 60 studies with 173 study cases that reported on the correlation between tree diversity and insect herbivore damage (53 study cases), abundance (52 study cases), incidence rate (40 study cases) and species richness (28 study cases), respectively. Regarding the metric of tree diversity, 94 and 7 study cases reported the Shannon or Simpson diversity of tree species, respectively, whereas 44 study cases reported the richness of tree species. In the remaining 28 study cases, we could apply a comparison between tree monocultures and mixtures of tree species, without any further quantification of tree species diversity.

In 92 study cases, herbivory was measured on a focal tree species whereas in the remaining 81 cases herbivory and especially herbivore abundance and species richness were measured at the plot level, either by summarizing herbivory over all tree species or by reporting capture rates in insect traps that could not be related to certain tree species. All study cases and study characteristics are documented in Appendix S6.4.

The included study cases covered a latitudinal range of -36.7° to 62.8° and spanned a gradient of -3.3°C to 26.9°C mean annual temperature (MAT). Study sites were most frequent in Europe and North to Middle America and sparse in Asia, Africa and Australia (Figure 6.2).

Pooled over all study cases that reported herbivore damage, abundance and incidence rate, the grand mean effect size was not significant and the corresponding funnel plot was symmetric (Table 6.1, Appendix S6.3, Egger's test: p = 0.58). There was a significant amount of residual heterogeneity indicating that heterogeneity in true effect-sizes could be accounted for by moderators.

Across the whole dataset (including effect sizes on herbivore species richness) the aspect of herbivory was a significant moderator of effect sizes as indicated by the significant omnibus test of model coefficients (Q_M = 10, df = 3, p = 0.02). When analysed separately, study cases on herbivore abundance and species richness yielded significantly positive grand mean effect sizes whereas herbivore damage and incidence rate showed no significant relationship with increasing tree diversity (Table 6.1, Figure 6.3). Funnels plots for these separate models were symmetric following Egger's regression test (damage: p = 0.14, abundance: p = 0.7, species richness: p = 0.25) except for herbivore incidence rate that displayed a positive correlation between effect sizes and the corresponding variance estimates (estimated effect sizes: 2.1 ± 0.9 standard error, p = 0.02).

Effect sizes for herbivore damage showed a decrease with MAT, indicating a tendency toward associational resistance in warm regions, and associational susceptibility in cold regions (Figure. 6.4), and were negative in experimental (estimated effect size: -0.02 ± 0.08, Table 6.1) and positive in observational studies (0.16 ± 0.11, Table 6.1). The abundance of generalist herbivores increased with tree diversity (estimated effect size: 0.15 ± 0.08, Table 6.1) while it decreases for specialist herbivores (-0.12 ± 0.1, Table 6.1). The included predictors explained

9.5 and 1.6 percent of variation for herbivore damage and abundance, respectively. As for herbivore incidence rate, effect sizes displayed a significantly positive relationship with MAT which accounted for 29.3 percent of variation. Effect sizes for herbivore species richness were not related to any of the proposed predictors (Table 6.1, Figure 6.4).

Addressing publication bias, most of funnel plots were symmetric, indicating that studies reporting positive and negative correlations, with low and high sample sizes were equally likely to be published (Appendix S6.3). The omnibus test of model coefficients indicated that effect sizes did not depend on the metric of tree diversity applied (Q_M = 5.66, df = 3, p = 0.129; Appendix S6.3). We found a negative correlation between the year of publication and MAT (r = -0.18, p = 0.01), showing that studies in cooler climates were carried out later, but not between the year of publication and the actual effect sizes (r = -0.01, p = 0.87), indicating that there was no temporal shift in our understanding of associational effects.

6.5 | Discussion

Our meta-analysis does not support previous claims of pervasive, globally consistent associational resistance (AR) to insect herbivores in mixed forests (Jactel & Brockerhoff 2007, Castagneyrol *et al.* 2014). Importantly, we show that several sources of variation in the strength and direction of associational effects have been overlooked, in particular mean annual temperature (MAT). We found no negative relationship between tree diversity and herbivory, regardless if the dataset was split according to the aspect of herbivory or the metric of tree diversity and, thus, have to reject our first hypothesis.

Instead, and in confirmation of our second hypothesis, we demonstrated that grand mean effect sizes differed between the four aspects of herbivory. Notably, herbivore species richness, an aspect not analysed in previous syntheses, increased with tree diversity. Importantly, we detected a significant relationship between MAT and the response of herbivore damage and incidence rate to increasing tree diversity. This finding partly confirms our third hypothesis and strongly improves our understanding of heterogeneity among studies.

We found that both aspects of herbivory displayed quite opposite tendencies. With increasing MAT, diversity effects on herbivore damage shifted from associational susceptibility (AS) in cold regions towards AR in warmer regions, whereas the opposite pattern was encountered for herbivore incidence rate, with a tendency towards AS in warmer regions.

As our study was based on correlation coefficients which become significant if any increase in tree diversity is accompanied by a steady increase or decrease in herbivory, we cannot make any statement on the change in the absolute amount of herbivory. However, we demonstrated that future syntheses on the strength of associational effects in forest systems should differentiate between the aspects of herbivory and consider the environmental context of the study site. This is critical because different aspects of tree diversity may influence damage (i.e. actual consumption) and incidence rate. For instance, incidence rate could reflect herbivore foraging behaviour, while damage additionally could include tree diversity effects on food quality and herbivore survival.

In search for factors that underlie the documented biogeographic gradient, we ask for more research on how associational effects depend on and follow global gradients in i) regional species diversity of herbivores and host plants, ii) herbivore density and pressure, especially during times of outbreaks (Coley & Barone 1996, Andrew & Hughes 2005, Kozlov 2008, Adams & Zhang 2009, Adams *et al.* 2009, Pennings *et al.* 2009, del-Val & Armesto 2010, Salazar & Marquis 2012, Kozlov *et al.* 2015), iii) herbivore specialization (Dyer *et al.* 2007, Jactel & Brockerhoff 2007, Schuldt *et al.* 2010, Castagneyrol *et al.* 2014, Forister *et al.* 2015, Schuldt *et al.* 2015), iv) tree defences (Coley & Barone 1996, Pennings *et al.* 2009, Pearse & Hipp 2012, Marquis *et al.* 2012, Salazer & Marquis 2012, Rodríguez-Castañeda 2013) and v) abiotic factors affecting tree or herbivore development (e.g. precipitation, climate stability, CO_2-concentration, UVB-radiation).

The literature on global gradients in herbivore pressure is controversial, providing examples of either increasing (Coley & Barone 1996, Kozlov 2008, Adams *et al.* 2009, Pennings *et al.* 2009, Salazar & Marquis 2012), unaffected (Andrew & Hughes 2005, Kozlov 2008, Pennings *et al.* 2009, Salazar & Marquis 2012) or decreasing (Adams & Zhang 2009, del-Val & Armesto 2010) pressure towards warmer climates. Regarding the level of background herbivory, e.g. the regular loss of woody plant foliage, a recent global analysis even pointed out that background herbivore pressure could show a nonlinear global pattern that is peaking in temperate regions (Kozlov *et al.* 2015).

The highest levels of herbivore pressure can be studied during insect outbreaks when herbivores face strong intraspecific competition and resource depletion. If, during outbreaks, herbivorous individuals are forced to forage less selectively, this might shift associational effects from AR to AS, regardless of the composition of tree species mixtures. Unfortunately, our dataset included only 5 study cases (3 original studies) that measured herbivory during outbreaks (S1 File) and we are not aware of any study that investigated global gradients in the frequency and intensity of insect herbivore outbreaks in forest ecosystems.

Global patterns in the relative abundance of generalist vs. specialist insect herbivores, however, had already been addressed. Here, Novotny *et al.* (2002) and Schuldt *et al.* (2010) proposed that tropical and subtropical forest are dominated by generalist herbivores, which, according to previous meta-analyses (Jactel & Brockerhoff 2007, Castagneyrol *et al.* 2014), are not affected or even benefit from the diversification of tree stands. A higher proportion of generalist species could explain the shift in herbivore incidence rates from AR to AS with increasing MAT. However, more comprehensive and global analyses concluded that the proportion of generalist species actually decreases towards tropical regions (Dyer *et al.* 2007, Forister *et al.* 2015).

Beside the difficulty to relate the documented shifts in associational effects with MAT to a single, underlying factor, any global synthesis of plant-plant-herbivore interactions might furthermore be confronted with gradients that are non-linear, non-additive or interrelated (i.e. show interactions such as latitudinal changes in plant defences being counterbalanced by higher herbivore pressure, Coley & Barone 1996, Pennings *et al.* 2009). In addition to the abiotic and biotic factors already mentioned, many decisions on the design of a study, such as the spatial scale of the investigated plant neighbourhood (Champagne *et al.* 2016), the age of the forest stand (Jeffries *et al.* 2006) and the sampling date (van Asch & Visser 2007) are likely to determine the sampling success of insect herbivore communities and thus impact the conclusions on the direction and strength of associational effects.

Given the multitude of potentially confounding factors, it is hardly surprising that our study documented a high amount of unexplained heterogeneity, and thus, highlights the limits of our meta-analytical approach. Here, research co-operations, such as the globally distributed network of tree diversity experiments (www.treedivnet.ugent.be), can offer great future opportunities to experimentally study associational effects along replicated global abiotic and biotic gradients (Verheyen *et al.* 2015).

6.6 | Conclusions

Our results indicate that studies of associational effects on herbivory might need to consider the biogeographical context in which plant-plant-herbivore interactions occur. Yet, along such gradients, joint impacts of insect herbivore diversity, pressure, specialization and abiotic factors on a global gradient of associational effects are difficult to disentangle.

We recommend that information on the damage, incidence, abundance and diversity, together with the identity of herbivores, be systematically recorded in observational and experimental tree diversity studies to provide a sounder understanding of mechanisms involved in AR and AS.

The inability to replicate major findings of previous meta-analyses, namely the significance of AR of mixed tree stands to insect herbivory (Jactel & Brockerhoff 2007, Castagneyrol *et al.* 2014), could even indicate that associational effects are nonlinear, thus adding another layer of complexity.

A better understanding of the mechanisms at play will require addressing the identity, functional characteristics, density and diversity of both, insect herbivores and focal tree species, in a systematic way along large geographical gradients.

Table 6.1 | Intercept-only and most parsimonious mixed-effects meta-regression models. Mixed-effects models tested the effect of mean annual temperature (MAT), study design (tree plantations or semi-natural forests) and herbivore specialization on the transformed correlation coefficient (Fisher's *z*-scores) between the diversity of tree species and the four different aspects of herbivory. In each model the intercept denotes the reference level of coefficient estimates, τ^2 denotes the variance between study cases, Q/Q_E relate to Cochran's Q-test for residual heterogeneity and Q_M denotes to the omnibus test of model coefficients. Significant parameter estimates are in bold.

Model	Model statistics			Parameters	Mean	Se	z-value	p-value
Grand mean all								
(= damage, abundance and	$\tau^2 = 0.11$			Intercept	0.01	0.03	0.31	0.76
incidence rate pooled)	$Q = 395.35$	df = 144	$p < 0.001$					
Damage								
Grand mean	$\tau^2 = 0.07$	AICc = 67.03		Intercept	0.06	0.06	1.12	0.26
	$Q = 92.24$	df = 52	$p < 0.001$					
Most parsimonious model	$\tau^2 = 0.07$	AICc = 65.63		Intercept (observational at zero MAT)	0.14	0.1	1.33	0.18
	$Q_E = 77.99$	df = 50	$p = 0.01$	**MAT**	**- 0.02**	**0.01**	**- 2.3**	**0.02**
	$Q_M = 8.22$	df = 2	$p = 0.02$	**Experimental vs. observational**	**0.23**	**0.1**	**2.24**	**0.03**
Abundance								
Grand mean	$\tau^2 = 0.19$	AICc = 93		**Intercept**	**0.08**	**0.04**	**2.13**	**0.03**
	$Q = 220.33$	df = 51	$p < 0.001$					
Most parsimonious model	$\tau^2 = 0.15$	AICc = 92.17		**Intercept**	**0.15**	**0.08**	**1.94**	**0.05**
	$Q_E = 210.51$	df=50	$p < 0.001$	**(generalist)**				
	$Q_M = 3.64$	df = 1	$p = 0.06$	**Specialists vs generalists**	**- 0.26**	**0.14**	**- 1.91**	**0.06**
Incidence rate								
Grand mean	$\tau^2 = 0.06$	AICc = 56.44		Intercept	- 0.08	0.07	- 1.08	0.28
	$Q= 73.18$	df = 39	$p < 0.001$					
Most parsimonious model	$\tau^2 = 0.04$	AICc = 43.54		**Intercept**	**- 0.42**	**0.07**	**- 5.67**	**< 0.001**
	$Q_E = 57.75$	df = 38	$p = 0.02$	**MAT**	**0.03**	**0.01**	**4.9**	**< 0.001**
	$Q_M = 23.97$	df = 1	$p < 0.001$					
Species richness								
Grand mean	$\tau^2 = 0.34$	AICc = 57		**Intercept**	**0.36**	**0.15**	**2.35**	**0.02**
(= most parsimonious model)	$Q = 99.23$	df = 27	$p < 0.001$					

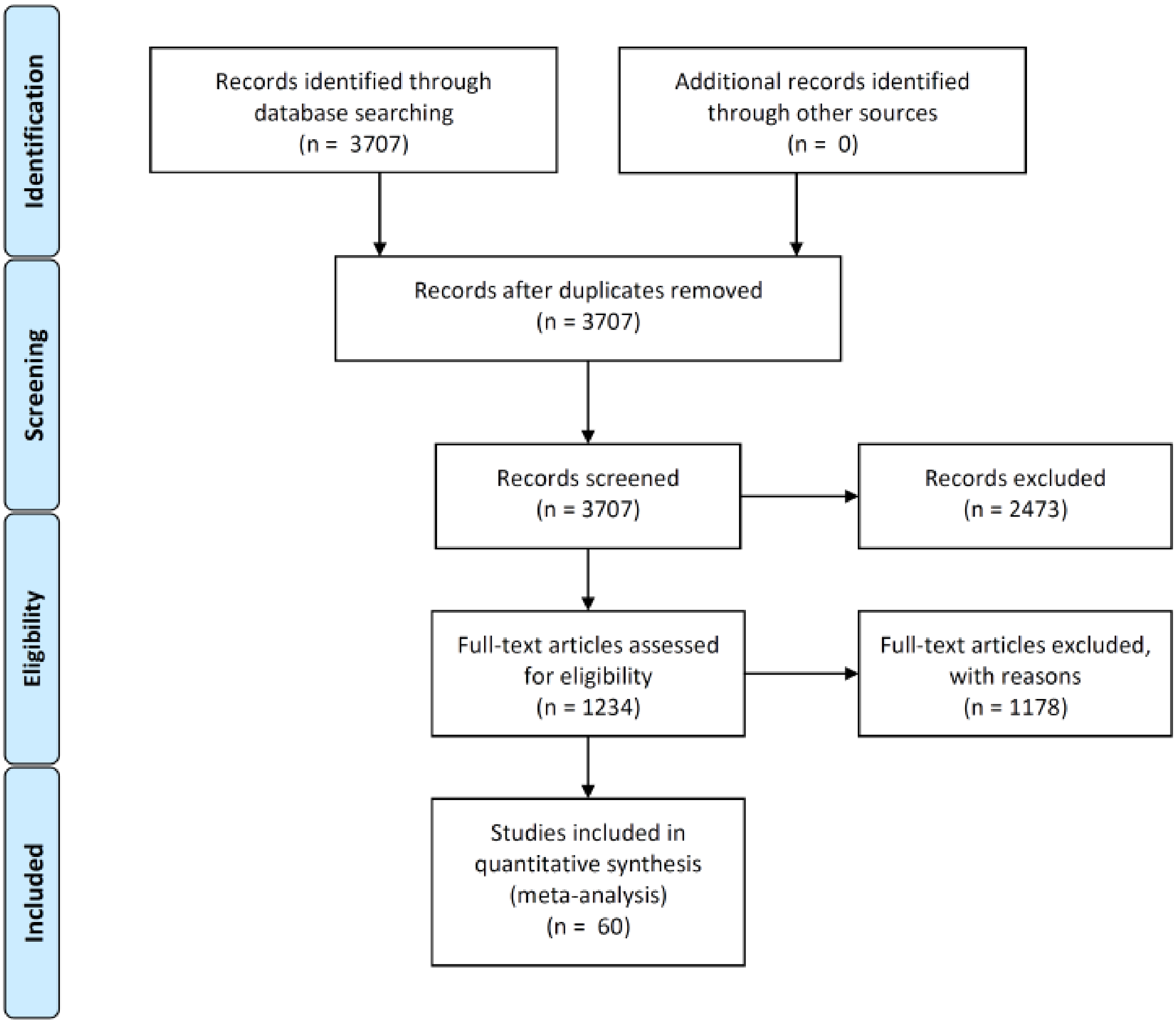


Figure 6.1 | Flow chart of the study search and selection process

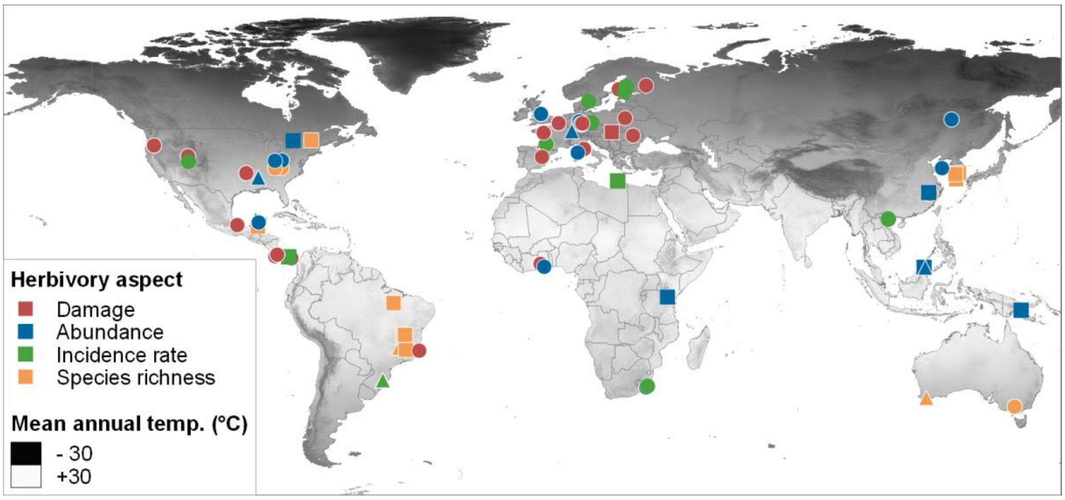


Figure 6.2 | Locations of the studies included in this meta-analysis reporting on the relationship between the four aspects of insect herbivory and tree Shannon/Simpson diversity (circles), tree species richness (squares) or the comparison between single vs. mixed stands (triangles). The colouring indicates the different aspects of insect herbivory. Made with Natural Earth. Free vector and raster map data (naturalearthdata.com).

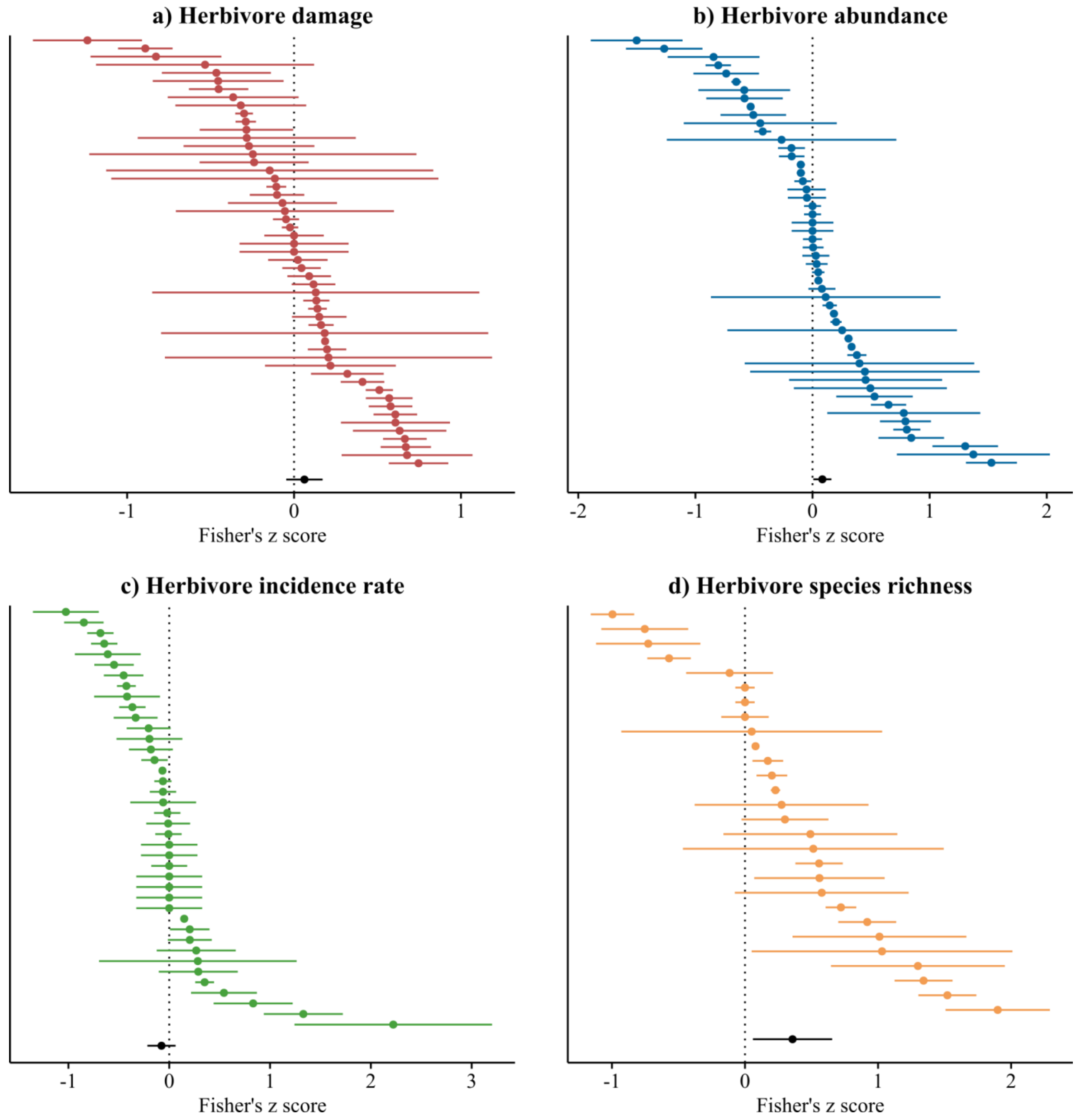


Figure 6.3 | Forest plots for the transformed correlation coefficients (Fisher's *z*-scores) between tree diversity and a) the damage sustained by, b) the abundance of, c) the incidence rate of, d) the species richness of insect herbivores. Each point represents the Fisher's *z*-score and the approximated confidence interval (= mean ± standard error × 1.96) for an individual study case. Negative values indicate associational resistance while positive values indicate associational susceptibility. Grand mean effect sizes, together with their 95% bootstrap confidence intervals, are shown in black at the bottom of each forest plot.

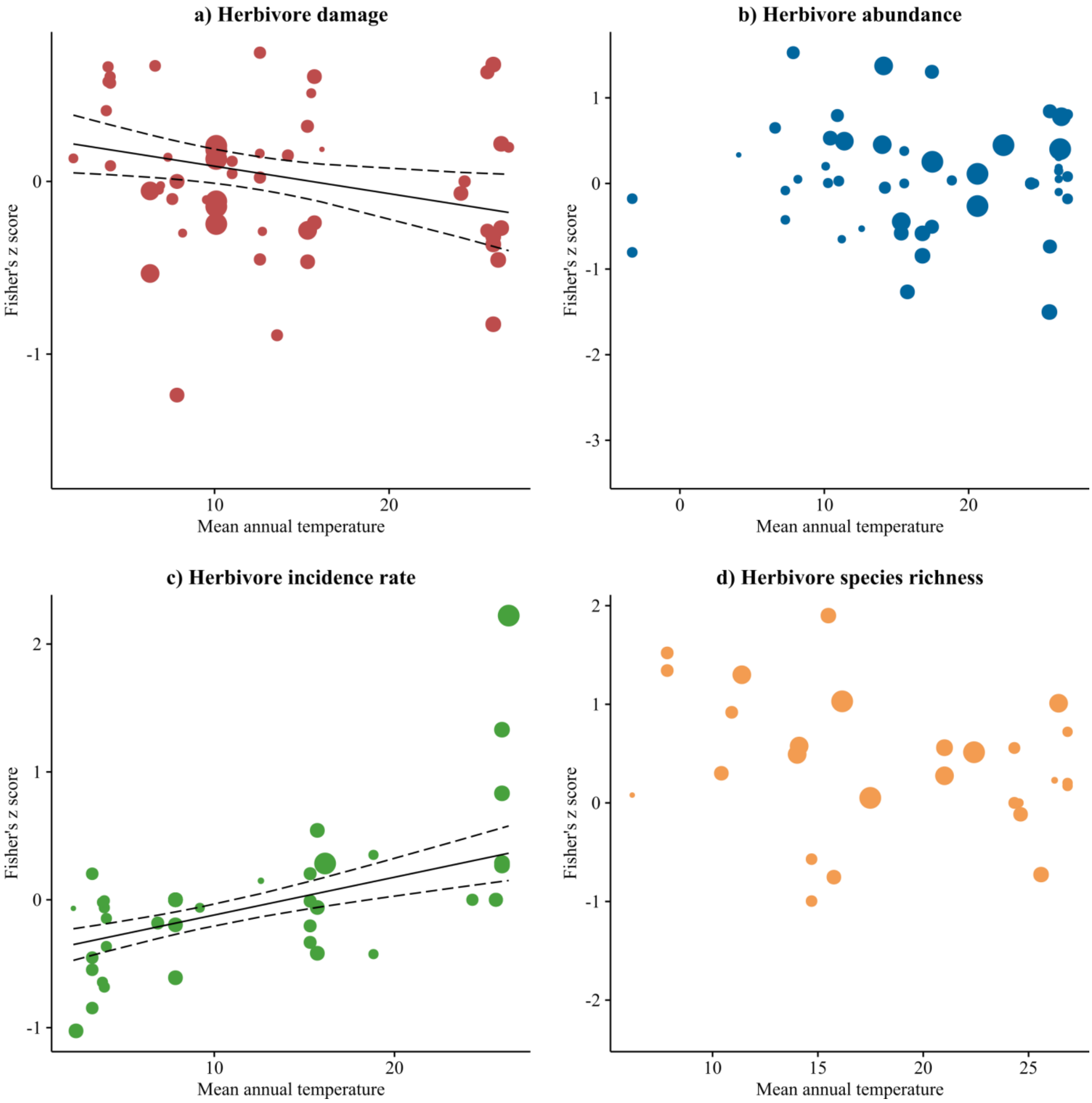


Figure 6.4 | Relationships between mean annual temperature (MAT) and the transformed correlation coefficients (Fisher's z-scores) between the diversity of tree species and a) the damage sustained by, b) the abundance of, c) the incidence rate of, d) the species richness of insect herbivores. Each point represents an individual study case for which negative values indicate associational resistance while positive values indicate associational susceptibility. The size of each point indicates its weight for estimating the regression slope (solid line) and the corresponding approximated credible interval (dotted lines) in a mixed-effects meta-regression model. Coefficient estimates for MAT are reported in Table 3.1.

6.7 | Author Contributions and Acknowledgements

S. Kambach and H. Bruelheide designed the study and drafted the manuscript together with the strong involvement of B. Castagneyrol and I. Kühn. S. Kambach performed the literature search and with the help of B. Castagneyrol and I. Kühn performed the data analysis. S. Kambach wrote the first draft of the manuscript, to which all co-authors contributed. The final manuscript is approved by all authors.

The work of S. Kambach was funded by the Helmholtz Centre for Environmental Research – UFZ, the Helmholtz Research School for Ecosystem Services under Changing Land-use and Climate (ESCALATE) and the Martin Luther University Halle-Wittenberg.

We are indebted to the researchers who shared their raw data with us. Publications and researches that provided data for this meta-analysis are listened in S1 File.

7 | Study 3 - Consequences of Multiple Imputation of Missing Standard Deviations and Sample Sizes in Meta-analysis

Stephan Kambach*[1,2,3], Helge Bruelheide[1,3], Jessica Gurevitch[4], Katharina Gerstner[3,5], Michael Beckmann[5], Ralf Seppelt[5,6]

* Stephan Kambach, Institute of Biology / Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Am Kirchtor 1, 06108 Halle, Germany, mail: stephan.kambach@gmail.com, phone: + 49 (0) 345 55 26263, fax: + 49 (0) 345 55 27228

1 Institute of Biology / Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, 06108 Halle, Germany

2 UFZ – Helmholtz Centre for Environmental Research, Department of Community Ecology, 06120 Halle, Germany

3 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

4 Department of Ecology and Evolution, Stony Brook University, Stony Brook, 11794-5245 New York, USA

5 UFZ – Helmholtz Centre for Environmental Research, Department Computational Landscape Ecology, 04318 Leipzig, Germany

6 Institute of Geoscience & Geography, Martin Luther University Halle-Wittenberg, 06099 Halle, Germany

Key-words Grand mean effect size, MICE, Multiple imputation by chained equations, Predictive mean matching, Random forest, Recommendation, Regression, Research synthesis, Simulated datasets, Unweighted analysis

7.1 | Abstract

1. Ecological meta-analyses often encounter incompletely reported studies that have missing variance measures (e.g. standard deviation values) or sample sizes that are necessary for incorporating the study in a weighted meta-analysis. Imputing these missing values can yield less biased meta-analysis results than the frequently adopted practice of omitting incompletely reporting studies or using unweighted analyses. Although previous publications already compared the performance of punctual imputation methods, no comprehensive analysis or recommendations for a larger set of imputation methods had been provided so far.

2. We conducted a systematic review on the frequency and treatment of missing variance measures in ecological meta-analyses. Based on simulated data sets that differed in size and spread, dependency and the proportion of missing standard deviations or sample sizes, we compared the performance of ten imputation methods that are provided by the R package mice. We then demonstrated how the imputation of missing standard deviations in a published dataset might alter the significance of the obtained grand mean effect size as well as the effect of covariates.

3. Our analyses revealed that neither the omission of incompletely reported studies nor the application of unweighted analyses is a viable solution to deal with missing variance measures in primary studies, although both have frequently been applied in meta-analyses in ecology and other disciplines. Most imputation methods yield less biased grand mean estimates than these approaches, but this varied with the characteristics of the data sets. Methods that are based on linear regression and predictive mean matching were particularly strong at accommodating correlations in the data sets and yielding unbiased grand mean estimates when missing standard deviations were imputed. These very same imputation methods resulted in biased grand mean estimates compared to all other imputation methods when effect size variances were based on sample sizes.

4. Multiple imputation of missing variance measures or sample sizes could help overcome the problem of incompletely reported primary studies in future meta-analyses, but caution must be exercised in consideration of potential correlations between the missing values and the chance to be missing.

7.2 | Introduction

Research synthesis aims at combining all available evidence on a research question to reach unbiased conclusions. In meta-analyses, individual effect sizes from different studies are summarized in order to obtain a grand mean effect size (hereafter "grand mean") and a corresponding confidence interval. Most of the analyses carried out in meta-analysis and meta-regression depend on inverse-variance weighting, in which individual effect sizes are weighted in order to accommodate differences in their precision and to separate within-study sampling error from among-study variation. Unfortunately, in original publications synthesized with meta-analyses in ecology and many other disciplines, missing and incompletely reported data are ubiquitous (Parker *et al.* 2016), especially for variance measures. Despite recent calls towards meta-analytical thinking and comprehensive reporting of variance measures (Hillebrand & Gurevitch 2013; Zuur & Ieno 2016; Gerstner *et al.* 2017), ecological meta-analyses may always face the issue of unreported variances, and certainly when older papers are incorporated in the synthesis.

In a review of 178 published ecological meta-analyses (cf. Supplement S7.1 and S7.2, Supporting Information), we found that nearly half of all studies (87) reported that the authors had encountered primary studies with missing albeit necessary variance measures (Fig. 7.1). Many of these meta-analyses (50) indicated that they only included those studies that completely reported all data needed to do the meta-analyses and excluded those studies that did not report the necessary variance measures. However, at the very least, excluding studies always means losing potentially valuable data. Moreover, when non-significant study results are more frequently missing variance measures and are therefore more frequently omitted from a meta-analysis than significant results, the estimated grand mean will be biased towards significance (a variant of the so-called "file-drawer problem", Rosenthal 1979; Idris & Robertson 2009). If these missing variances were imputed, this could potentially lead to less biased grand mean estimates compared to the exclusion of incompletely reported effect sizes (Furukawa *et al.* 2006; Nakagawa & Hauber 2011; Idris 2011; Idris, Abdullah & Tolos 2013; Nakagawa 2015; Ellington *et al.* 2015). Besides, imputing missing variances would contribute to an increase in the number of effect sizes and thereby the precision of the obtained estimates of the grand mean (Idris & Robertson 2009) or subgroup means. Moreover, the larger dataset might also permit one to test hypotheses that were not testable with only the subset of completely reported effect sizes (e.g. on the factors that account for differences in effect sizes). However, missing variances have rarely been imputed in ecological meta-analyses (Fig. 7.1).

A range of techniques and statistical models can be applied to impute missing variances but their reliability and accuracy has not been well investigated. In order to be able to provide general recommendations on the imputation method of choice for ecological data, we need to assess the effects of a number of imputation methods on the estimated grand means of artificial data sets with specific simulated properties. To our knowledge, all previous studies that compared the effects of different imputation methods were done on published data sets and with a limited number of imputation methods (e.g. see Robertson, Idris & Boyle 2004; Furukawa *et al.* 2006; Wiebe *et al.* 2006; Thiessen Philbrook, Barrowman & Garg 2007; Idris & Robertson 2009; Idris, Abdullah & Tolos 2013; Ellington *et al.* 2015).

To impute missing values, van Buuren (2012) recommended two techniques, expectation maximization (EM) and Markov chain Monte Carlo (MCMC) algorithms. Both can yield unbiased parameter estimates under the assumption that values are missing at random. In this study we focussed on an MCMC method, also termed multivariate imputation by chained equations (MICE), because EM algorithms assume multivariate normality of all variables, a prerequisite that constrains applicability, especially in ecological data sets. The MICE algorithm, in contrast to joint modelling, allows a fully conditional specification, i.e. a separate imputation model for each variable that has missing values (van Buuren 2012). Furthermore, the suitability of the MICE algorithm to impute missing variability estimates in meta-analysis data sets has previously been demonstrated in both medical meta-analyses (Resche-Rigon *et al.* 2013; Jolani *et al.* 2015; Quartagno & Carpenter 2016) and ecological meta-analyses (Ellington *et al.* 2015). In *R* (R Development Core Team 2013) there are currently two frequently used packages to run MICE algorithms, namely *mi* (Su *et al.* 2011) and *mice* (van Buuren & Groothuis-Oudshoorn 2011). We used the *mice* package because it already includes a range of easily applicable imputation methods.

In this study, we explore how ten methods to impute missing variance measures (here, missing SDs) differ in their effect on the grand mean estimates under the consideration of varying properties of the meta-analysis dataset, namely (see also Table 7.1):

1. the dataset size and proportion of missing SDs,
2. the spread of SDs,
3. the correlation between SD and effect size or covariate values and
4. the correlation between the probability of missing SDs and effect size or size or SD values.

If SDs are missing, but the numbers of samples are reported, effect sizes variances can be approximated using studies' sample sizes alone (Hedges & Olkin 2014, Koricheva, Gurevitch & Mengersen 2013). Since the performance of imputation methods might vary depending on whether missing SDs or missing sample sizes are imputed, we repeated the outlined deletion-and-imputation approach for studies' sample sizes instead of SDs and weighted the studies by the inverse of the approximated variance measures.

Based on these findings, we then demonstrate the applicability of multiple imputation of missing SDs with re-analysing the published and open-access dataset of Marczak, Thompson & Richardson (2007).

7.3 | Materials and Methods

Creation of artificial meta-analysis data sets with simulated data patterns

We simulated meta-analysis data sets in which each row corresponds to a study case (i.e. effect size) that compared a measurement between a control and a treatment group. Effects were quantified with log response ratios and the corresponding variance estimates (eqn 7.1, eqn 7.2) based on the measurements' means (x_t, x_c), standard deviations (sd_t, sd_c), and sample sizes (n_t, n_c) in the treatment and control group. If SDs are not reported, an approximated variance measure can be calculated from the sample sizes alone (eqn 7.3, Hedges & Olkin 2014).

eqn 7.1 $$logRR = \log(\frac{x_t}{x_c})$$

eqn 7.2 $$var(logRR) = \frac{sd_t^2}{n_t x_t^2} + \frac{sd_c^2}{n_c x_c^2}$$

eqn 7.3 $$var(logRR)_{approx} = \frac{n_t + n_c}{n_t * n_c}$$

Ecological meta-analyses often summarize results across different study characteristics (e.g. different biomes, experimental set-ups, species groups and environmental gradients). Those different categorical and numerical covariates would be expected to result in different effect sizes and SDs and are thus usually modelled in ecological meta-analysis data sets to account for heterogeneity in study outcomes. In order to simulate the influence of such categorical and continuous covariates on the effect size SDs, we first randomly assigned an integer covariate (values ranging from 1 to the number of effect sizes in the respective dataset) and a categorical covariate (with two categories) to each row (i.e. each study case) of the simulated dataset. In all studies we set the mean value for the control group to one, the SD to 1 and the sample size to 10. For the treatment group we spread the mean values between a minimum of $e^{\log(2) - 0.5 \times \log(2)}$ and a maximum of $e^{\log(2) + 0.5 \times \log(2)}$. The SD of the treatment group was then set as the treatment mean divided by two. These steps were applied in order to derive a "base" dataset with the following properties (illustrated in Fig. 7.2). (i) Effect sizes (i.e. response ratios between treatment and control groups) were evenly spread between the minimum and maximum value. (ii) The variance estimates of effect sizes (eqn 7.2) had the same value. (iii) We applied a linear mixed-effects model that included the identity of each study case as a random term in order to estimate the grand mean of the dataset. The true grand mean was exactly log(2), indicating that measurement values of the treatment groups were, on average, twice as high as the measurement values of the control groups.

This "base" dataset was subsequently modified to create a total of 216 simulated data sets that differed in the following characteristics (described in more detail in Table 7.1): dataset size, spread of SDs, correlation of SDs and correlation of SD missingness. In each dataset we simulated missingness by deleting between 5% and 95% of the SDs in the treatment groups (in steps of 5%) with the chance of each SD to be deleted according to the characteristics of the artificial data sets as outlined in Table 7.1.

To test the effects of missing sample sizes on the grand mean estimates, we additionally constructed 216 simulated data sets in which we deleted sample sizes and that varied in the following dataset characteristics: the dataset size, the spread of sample sizes, the correlation of sample sizes (in contrast to the SD data sets, we created a positive correlation between sample size and effect sizes values so that higher effect sizes likewise gained smaller effect size variances) and the correlation of sample size missingness (in contrast to the SD data sets, we created a negative correlation between sample sizes and their chance of being deleted, simulating that studies with a lower sampling effort would more frequently omit this information).

Imputation of missing SDs

In the reduced data sets all deleted SDs were filled using the multiple imputation algorithm that is implemented in the *mice* package in *R* (Fig. 7.3, van Buuren & Groothuis-Oudshoorn 2011; R Development Core Team 2013). Within each dataset, the imputations were based on all available variables (control and treatment mean values, SD and sample sizes as well as the numerical and categorical covariate). For each dataset, deleted SDs were imputed using ten imputation methods (cf. Table 7.2): the omission of effect sizes with missing SDs (*deletion*), mean value imputation (*mean*), random sampling from available SDs (*sample*), linear models (*linear.bayes*, *linear.reg* and *linear.pred*, *pmm*), random forest (*rf*), classification and regression trees (*cart*) and the application of an unweighted meta-analysis (*unweighted*).

To account for the inherent uncertainty of imputed SDs every imputation was repeated five times. For each of the five imputed data sets we estimated the data sets' grand mean with a linear intercept-only mixed-effects model that weighted effect sizes by the inverse of their corresponding variance estimate and treated the identity of each study case as a random effect (*rma* function in the *metafor* package, Viechtbauer 2010). The resulting five grand mean estimates as well as their corresponding approximated 95% confidence intervals were averaged with the *pool* function of the *mice* package (illustrated in Fig. 7.3) and compared to the grand mean of the full dataset (i.e. without deletions), both visually and via calculation of the root mean square error (RMSE).

Imputation of missing sample sizes when effect sizes were an approximated variance measure

For the imputation of missing sample sizes, we applied the exact same workflow as we applied for the imputation of missing SDs but on the data sets in which we previously deleted the sample sizes of the treatment group.

All analyses were conducted in *R* (R Development Core Team 2013) using the packages *data.table* for data wrangling (Dowle *et al.* 2015), *ggplot2* and *cowplot* for graphical representations (Wickham 2009; Wilke 2016), *metafor* for meta-analytical models (version 1.9-9, Viechtbauer 2010) and *mice* for multiple imputations (version 2.30, van Buuren & Groothuis-Oudshoorn 2011).

7.4 | Results

Consideration 1: Dataset size and proportion of missing SDs

The larger the dataset and the lower the proportion of missing SDs, the more accurately missing SDs could imputed, regardless of the method applied (cf. Supplement S7.3, Supporting Information). When the dataset included 50 or more study cases and the spread of SDs was low (with values between 0.1 and 1), all imputation methods yielded grand mean estimates that were nearly identical to the true grand mean of the full data set, regardless of the proportion of missing SDs (Fig. S7.3). If the spread of SD values was large (with values between 0.1 and 100), however, all imputation methods tended to yield a higher variability and thus potential bias in the grand means above a threshold of 50% of missing SDs.

Linear regression-based imputation methods (methods *linear.bayes*, *linear.reg*, *linear.pred*) could potentially predict negative SDs. Up to 30% of the SDs that were imputed with linear regression were negative when either the dataset size was small (10 or 20 studies) and less than 8 SDs were reported or when the that dataset size was large (50, 100 and 200 studies) but less than 20 SDs were reported. Effect sizes corresponding to these negative SDs had to be omitted from the subsequent estimation of the grand mean (which is not apparent from the presented figures).

All imputation techniques that relied on predictors to impute missing SDs (all methods beside *mean* and *sample*) must be supplied with at least, between three and six reported SDs to converge and yield imputation results. In the case of the randomly distributed and missing SDs, the omission of incompletely reported effect sizes (method *deletion*) yielded unbiased grand mean estimates but overly large confidence intervals that even more increased with the number of missing SDs (Fig. 7.4 and Fig. S7.3). The application of unweighted analyses (method *unweighted*) likewise resulted in oversized confidence intervals (Fig. S7.3). In this scenario of random distribution and deletion of SDs, all imputation methods yielded unbiased grand mean estimates and confidence intervals, with the only exception of *linear.reg* leading to oversized confidence intervals when the dataset was small and had a high proportion of missing SDs.

Consideration 2: Spread of SDs

Increasing the spread of SDs led to an increase in the RMSE of estimated grand means in comparison to the grand mean of the full dataset, regardless of the imputation method (Fig. 7.5). Unweighted analyses resulted in the largest RMSE whereas all other methods yielded comparable RMSEs that did not follow any ranking of imputation methods. An increase in SD spread increased the leverage of effect sizes with small SDs on the grand mean estimate and decreased the leverage of effect sizes with large SDs. Thus, the impact of “misclassifying” deleted SDs (with regard to their respective values in the full dataset) on the grand mean estimate of the imputed dataset increased with the spread of the SDs.

Consideration 3: Correlation of SDs with effect sizes or covariates

When SDs correlated with effect sizes an unweighted analysis resulted in biased grand mean estimates that significantly deviated from the estimate of the full dataset (see Fig. 7.4, demonstrated on a dataset of 500 study cases with SDs ranging from 0.1 to 100). As the proportion of deleted SDs increased, most imputation methods yielded grand mean estimates that gradually approximated the estimate from an unweighted analysis. In case of the *sample* and *rf* method this convergence occurred linearly. In case of *deletion*, *mean*, *pmm* and *cart* methods, grand mean estimates remained very similar to the full dataset up to a threshold of ca. 70-80% of missing SDs. Above that ratio of missing SDs, the estimated grand means rapidly approximated the (biased) grand mean of an unweighted analysis. Imputations based on linear models (*linear.bayes*, *linear.reg* and *linear.pred*) always yielded unbiased grand mean estimates and confidence intervals regardless of the proportion of deleted SDs.

Consideration 4: Correlation of SD missingness with effect size or size of SDs

When the chance of deleting SDs negatively correlated with the effect size, this did not bias the imputed grand mean estimates and confidence intervals (Fig. 7.4). However, when the chance of deleting SDs correlated with the size of SDs, this led to a gradually biased underestimation of confidence intervals as the proportion of deleted SDs increased (Fig. 7.4). In the cases of *sample*, *rf* and *cart* methods, this confidence interval shrinking happened roughly linearly, whereas in the cases of *linear.bayes*, *linear.reg*, *linear.pred*, *mean* and *pmm* methods, it started above a threshold of between 50 and 80% of deleted SDs. The omission of study cases with missing SDs (method *deletion*) was the only method that yielded unbiased grand mean estimates and confidence intervals, regardless of the proportion of deleted SDs.

Consideration 4: Correlation of SD missingness with effect size or size of SDs

In the case of SDs being correlated with effect sizes and their chance of deletion correlated with their value, unweighted analyses yielded biased grand mean estimates that were outside of the confidence interval of the full dataset (Fig. 7.6). With increasing proportion of deleted SDs, the imputation methods of *sample*, *mean*, *pmm*, *rf* and *cart* all resulted in grand mean estimates that gradually converged to the estimate that would be obtained by an unweighted analysis. This convergence was linear in the case of *sample* and *rf* imputation methods, whereas in the case of *mean*, *pmm* and *cart* methods this convergence started above a threshold of between 50 and 70% of deleted SDs. The omission effect sizes with missing SDs (method *deletion*) as well as imputation methods that were based on linear regressions (*linear.bayes*, *linear.reg* and *linear.pred*) all yielded unbiased grand mean estimates and corresponding confidence intervals (Fig. 7.6).

Imputation of missing sample sizes

If effect sizes were weighted by an approximate variance estimate that only used sample sizes (eqn 7.3) all imputation methods yielded unbiased grand mean estimates (S7.4). These estimates were generally closer to the grand mean of the full dataset than in the case of the imputation and weighting via SDs (cf. Fig. 7.5 and Fig. S7.4). However, if the chance of deleting sample sizes correlated with their value while this value also correlated with the corresponding effect size, those imputation methods that were based on linear regression (method *linear.bayes*, *linear.reg* and *linear.pred*) yielded grand means estimates that were biased and linearly related the proportion of deleted sample sizes (Fig. 7.6). All other imputation methods (method *sample*, *mean*, *pmm*, *rf* and *cart*) still yielded unbiased grand means and confidence.

Imputation of missing SDs in a published ecological meta-analysis dataset

The meta-analysis of Marczak, Thompson & Richardson (2007) summarized the response of consumer biomass and density to resource subsidies across different recipient habitat types, trophic levels and functional groups. Due to missing SDs, the authors applied an unweighted analysis, but unlike most studies, they still provided all available SDs (for 94 out of 116, i.e. 80% of the study cases). After checking that mean effect sizes with available and missing SDs did not differ significantly (Student's t-test = 0.7, df = 18.3, p = 0.5) we decided to impute the missing SDs via predictive mean matching (method *pmm*). The details of our imputations and re-analyses are given in S7.5. Due to the uncertainty introduced through multiple imputations the estimated grand mean of the imputed dataset became non-significant in our re-analysis (Fig. 7.7). In contrast, habitat-level aggregated mean effect sizes showed mostly narrower confidence intervals, and thus, higher significance levels as compared to the unweighted analysis.

7.5 | Discussion

Missing variance measures and sample sizes are a prevalent problem in research synthesis. Yet, few ecological meta-analyses have adapted imputation algorithms to handle missing values (Fig. 7.1). Our study shows that the use of unweighted analyses and the omission of incompletely reported studies can both lead to biased estimates of the grand mean and oversized confidence intervals. While imputations generally yielded more accurate results than the analyses of unweighted or incomplete data sets, none of our selected imputation methods emerged as the single best method under all dataset characteristics (Table 7.2). Thus, for any meta-analysis dataset the method of choice to impute missing variances (here, missing SDs) or sample sizes must be chosen under the following considerations:

1. Dataset size and proportion of missing SDs - When the dataset is large or only a small proportion of SDs are missing, all imputation methods can yield unbiased grand mean estimates and confidence intervals. Moreover, all of the imputation methods outperformed the practices of omitting incompletely reported studies and the application of unweighted analyses. As the proportion of missing SDs increased, the chance of linear regression-based imputation methods (*linear.bayes*, *linear.reg* and *linear.pred*) to estimate negative SDs (which must be omitted in subsequent analyses) increased, leading to overly large confidence intervals. When the proportion of missing SDs is large, *rf* imputations might also fail to converge and thus be unable to impute missing SDs. In this case, imputation via predictive mean matching (*pmm*) would still result in plausible SD estimates.

2. Spread of SDs - As the spread of SDs increased, the impact of “misclassifying” SDs (with respect to the full dataset) likewise increased and differences between imputation methods were amplified. Thus, in ecological meta-analysis data sets that could cover very heterogeneous study designs and SDs, the effect of using different imputation methods might be stronger than in clinical meta-analyses across well-replicated studies. In those cases of heterogeneous SDs, researchers might want to apply different imputation techniques in order to assess the reliability of the obtained grand mean estimates.

3. Correlation of SDs with effect sizes or covariates - With an increasing proportion of missing SDs, all imputation methods, except those that are based on linear regressions (*linear.bayes*, *linear.reg* and *linear.pred*), converged towards the biased grand mean estimate of an unweighted analysis. In cases where linear regression based techniques cannot be applied, imputation via *mean* value, *pmm* and *cart* methods can yield unbiased grand mean estimates up to a threshold of ~ 70% of missing SDs.

4. Correlation of SD missingness with effect size or size of SDs - When the largest SDs are those that are most often missing, grand mean estimates will consequently have too narrow confidence intervals. In our simulations, both imputation methods *linear.pred* and *mean* yielded accurate confidence intervals up to a higher threshold of missing SDs in contrast to the rest of the tested imputation methods. However, when incompletely reported effect sizes would anyway have the lowest leverage on the estimated grand mean (because they have the highest SDs), their omission would yield unbiased grand mean estimates and confidence intervals.

5. Imputation of missing sample sizes - Most imputation methods could yield unbiased grand means regardless of the dataset characteristics and the proportion of missing SDs. Similar to the imputation of missing SDs, omitting incompletely reported effect sizes or applying unweighted analyses mostly resulted in oversized confidence intervals. However, if studies that exhibit larger effect sizes have larger sample sizes and studies of smaller sample sizes tend to not report these values then the tested imputation methods via linear models (method *linear.bayes*, *linear.reg* and *linear.pred*) cannot accommodate this hidden data structure and, as a result, can lead to strongly biased grand mean estimates. Should researchers suspect this peculiar pattern in their dataset, we recommend relying on other methods like predictive mean matching, random forest or classification and regression trees.

When deciding on the best method to handle missing SDs and sample sizes in meta-analyses, researchers need to balance all four dataset considerations and decide on dataset characteristics that most importantly determines which, if any, imputation method can yield reliable grand mean estimates and confidence intervals. Patterns of data missingness might be assessed by checking if large effect sizes, small SDs or large sample sizes are suspiciously over-represented in the dataset at hand. These checks can be conducted e.g. via funnel plots, Egger's regression tests on funnel plot asymmetry (Sterne & Egger 2005) and comparisons of mean effect sizes between completely and incompletely reported study results. Correlations between SDs or sample sizes and study features could be tested via correlation and regression analyses.

As tools for the multiple imputation of missing variance measures have become readily available only recently, we expect them to become standards in future meta-analyses. Our re-analysis of the Marczak, Thompson & Richardson (2007) meta-analysis exemplifies this approach, and shows that multiple imputation might, in some cases, even change the conclusions derived from a meta-analysis.

With our simulation study, we aimed to raise more awareness on the problem of incompletely reported study results (Parker *et al.* 2016; Gerstner *et al.* 2017) and their imputation. In *R*, a handful of packages, beside *mice* (van Buuren & Groothuis-Oudshoorn 2011), are capable to run multiple imputation algorithms and a comprehensive overview of the most prominent packages can be found in Nakagawa & Freckleton (2011).

We discourage meta-analysts from using unweighted analyses because these are most likely to result in biased grand mean estimates. If, however, data sets cannot provide valid predictors for the imputation of missing SDs or sample sizes, simply filling all missing values with the mean or median (in the case that SDs are not normally distributed) of all reported values would still allow the inclusion of incompletely reported effect sizes. With this approach, effect sizes with small SD or large sample sizes will still have the highest impact on the obtained grand mean estimate. In our opinion, this option should always be preferred to an unweighted analysis that neglects all differences in effect size reliability. Weighting studies by the approximated variance estimate that is solely based on sample sizes should, again in our opinion, only be applied if almost no SDs are available because this variance estimate comprises less information and is only a good variance proxy if effect sizes and missing SDs are normally distributed, a prerequisite that is might be rarely fulfilled in ecological data sets.

In summary, our study provides evidence that future meta-analysts would benefit from routinely applying imputation algorithms to fill unreported study variances (e.g. SDs and sample sizes) in their data sets to increase both, the inclusion of all available data and the validity of derived conclusions.

7.6 | Author Contributions and Acknowledgements

JG, KG, MB, RS, SK and HB conceived the ideas; SK reviewed the published meta-analyses, performed the simulations and re-analysis and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

This work was supported by the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation (DBI-1052875 and 1119891) and by the Helmholtz Centre for Environmental Research (Research School ESCALATE, VH-KO-613), the German Federal Ministry of Education and Research (GLUES, 01LL0901A) and the Synthesis Centre (sDiv) of the German Centre for Integrative Biodiversity Research (DFG FZT 118).

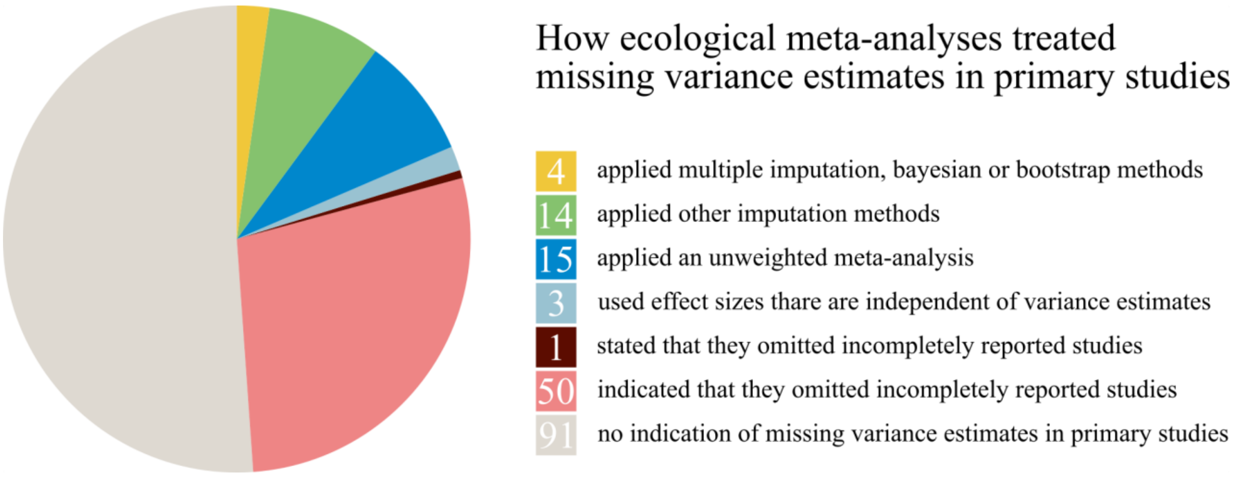


Figure 7.1 | Treatment of missing variance measures in primary studies by 178 ecological meta-analyses published until September, 2015 (cf. Supplement S7.1 and S7.2). Numbers in coloured boxes refer to the number of meta-analyses in the respective group.

Table 7.1 | Characteristics of the artificial meta-analysis datasets. SD: standard deviation.

Dataset characteristic	Implication for the calculation of the grand mean	Treatment in artificial datasets
1. Dataset size and proportion of missing SDs	With an increasing number of study cases the precision of the grand mean estimate increases (i.e. the according confidence interval becomes smaller)	Dataset size was set to 10, 20, 50, 100, 200 or 500 studies
2. Spread of SD values	The larger the spread of SDs, the stronger the impact of the smallest SDs and the weaker the impact of the highest SDs on the estimation of the grand mean	SDs were evenly spread between 0.1 and 1, 0.1 and 10 or 0.1 and 100
3. SDs correlate with effect sizes or covariate values	Studies with certain features (e.g. large effect size or experimental vs. observational set-up) always exhibit a high impact on the estimated grand mean	SDs were assigned at random or correlated with $\frac{1}{(treatment\ mean)4}$, a numeric or a categorical covariate (Fig. 2)
4. Correlation of SD missingness with effect size or size of SDs	If studies with certain features (e.g. small effect sizes or large SDs) are frequently omitted this results in a biased grand mean estimate that does not summarize all available studies	The probability that individual SDs were deleted was uniform or increased with or $\frac{1}{(treatment\ mean)4}$ or SD^4

Table 7.2 | Methods applied to impute deleted standard deviations (SDs) in artificial meta-analysis datasets. All methods are implemented in the *mice* package and described in more detail in van Buuren & Groothuis-Oudshoorn (2011) and van Buuren (2012). β_0: intercept, β_1: slope, ε: error term.

Imputation method	Abbreviation	Description of the method
List wise deletion	*deletion*	Omits all studies with missing SDs from the calculation of the grand mean.
Unweighted analysis	*unweighted*	Calculates the grand mean without weighting effect sizes differentially.
Bayesian posterior	*linear.bayes*	Fits a Bayesian linear regression model: $SD = \beta_0 + \beta_1 + \varepsilon$. Missing values are imputed by drawing β_0, β_1 and ε from their respective posterior distributions.
Linear prediction with model error	*linear.reg*	Fits a linear regression model: $SD = \beta_0 + \beta_1 + \varepsilon$. Missing values are imputed with the respective estimates of β_0, β_1 and a stochastic error according to ε.
Linear prediction without model error	*linear.pred*	Fits a linear regression model: $SD = \beta_0 + \beta_1 + \varepsilon$. Missing values are imputed with the respective estimates of β_0 *and* β_1.
Random sample	*sample*	Missing values are imputed by drawing a random SD value from all the reported SDs.
Mean value	*mean*	Missing SDs are all filled with the mean value of all reported SDs.
Predictive mean matching	*pmm*	Fits a linear regression model: $SD = \beta_0 + \beta_1 + \varepsilon$. Missing values are imputed by randomly drawing one of the three reported SDs whose predicted values (by the respective estimates of β_0 *and* β_1) are the closest to the predicted value of the missing SD.
Random forest	*rf*	Implementation of Breiman's random forest algorithm (Breiman 2001). Missing values are imputed with the average of 10 predictions from classification and regression trees that are based on random subsets of predictor variables.
Classification and regression tree	*cart*	Machine learning algorithm that seeks cutting points in the set of predictor values that divide the data into homogeneous subsamples. Missing values are imputed by randomly selecting of one reported SD from the subset of data that is classified in the same terminal tree node as the missing SD.

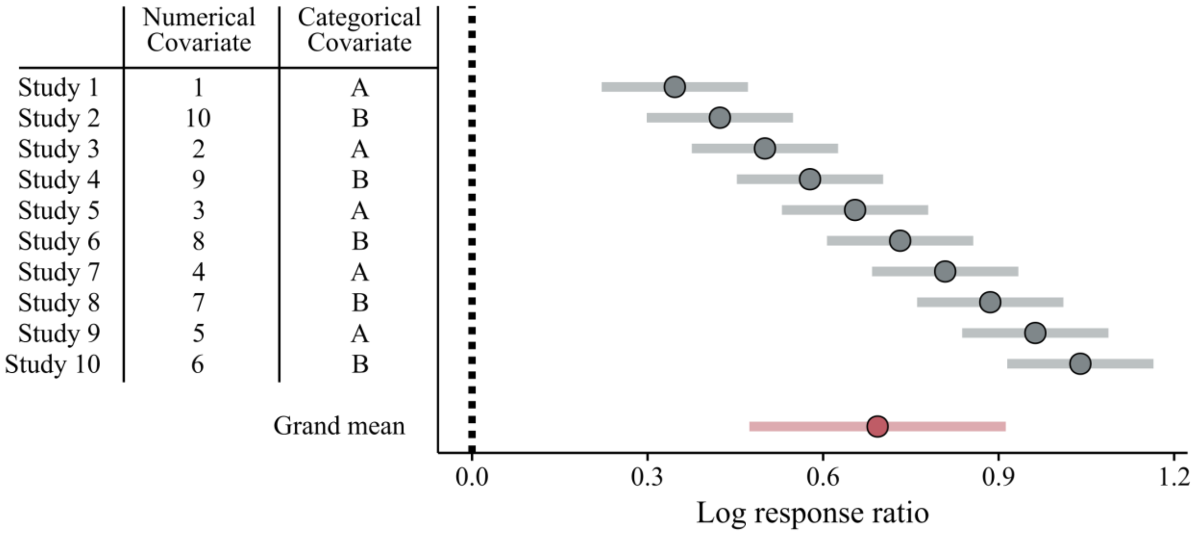


Figure 7.2 | A forest plot of an artificial meta-analysis dataset with ten study cases. Points denote individual effect sizes (i.e. log response ratios between the mean of the treatment and control groups, eqn. 1). Horizontal lines indicate the corresponding sampling variance estimates (eqn. 7.2). The grand mean together with the corresponding approximated 95% confidence interval, as calculated from a linear mixed-effects model, are depicted at the bottom.

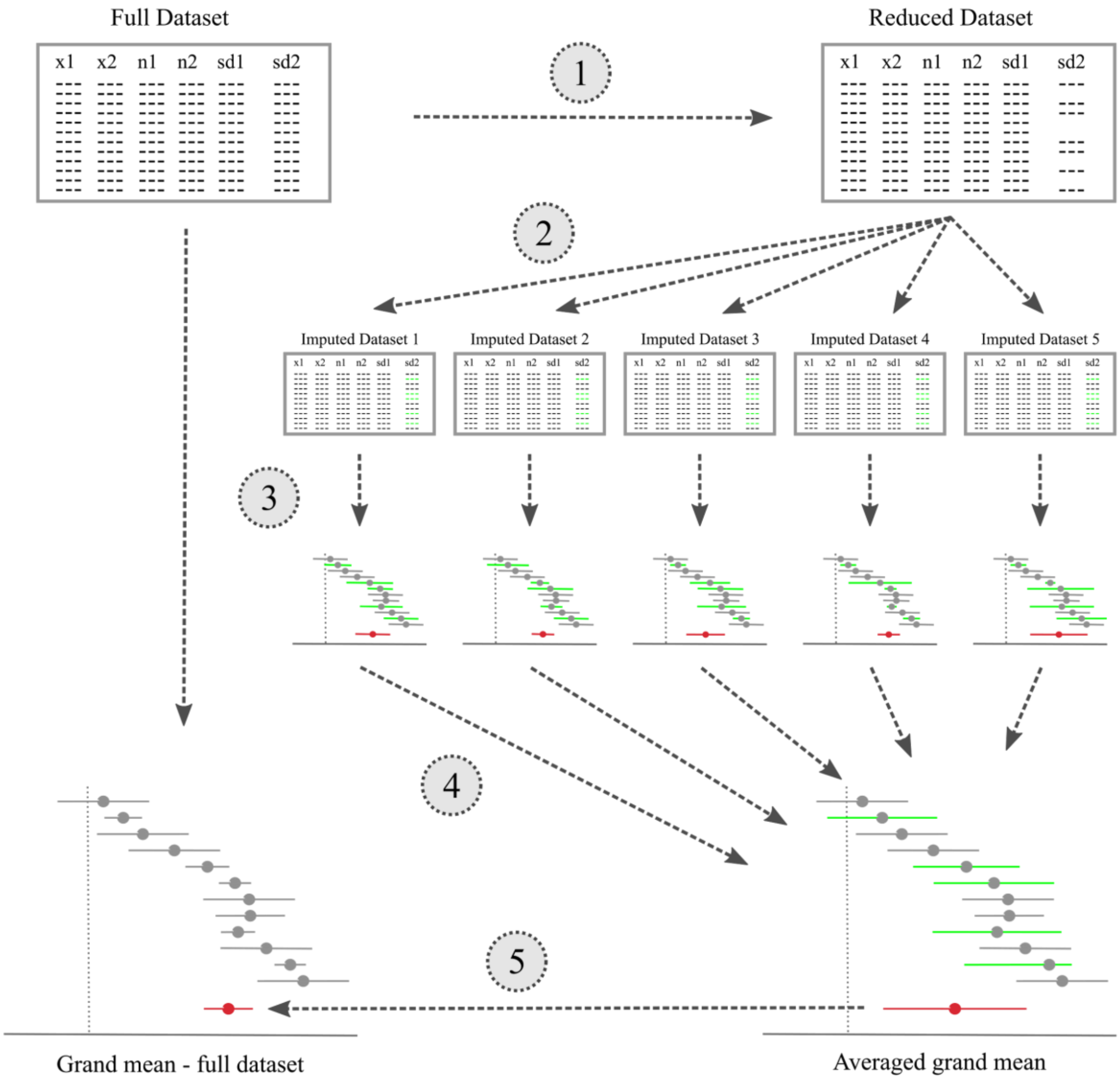


Figure 7.3 | Workflow for the multiple imputation of deleted standard deviations (SDs). (1) Deletion of 5% up to 95% of the SDs in treatment groups of an artificial dataset. (2) Deleted SDs are imputed (in green) via multiple imputations (five times), all done with the same imputation method. (3) Separate mixed-effects meta-analyses on each of imputed dataset (imputed SDs in green). (4) Averaging of the five grand means and their corresponding confidence intervals. (5) Comparison of the average imputed grand mean and confidence interval with the estimates from the full dataset (i.e. without missing SDs).

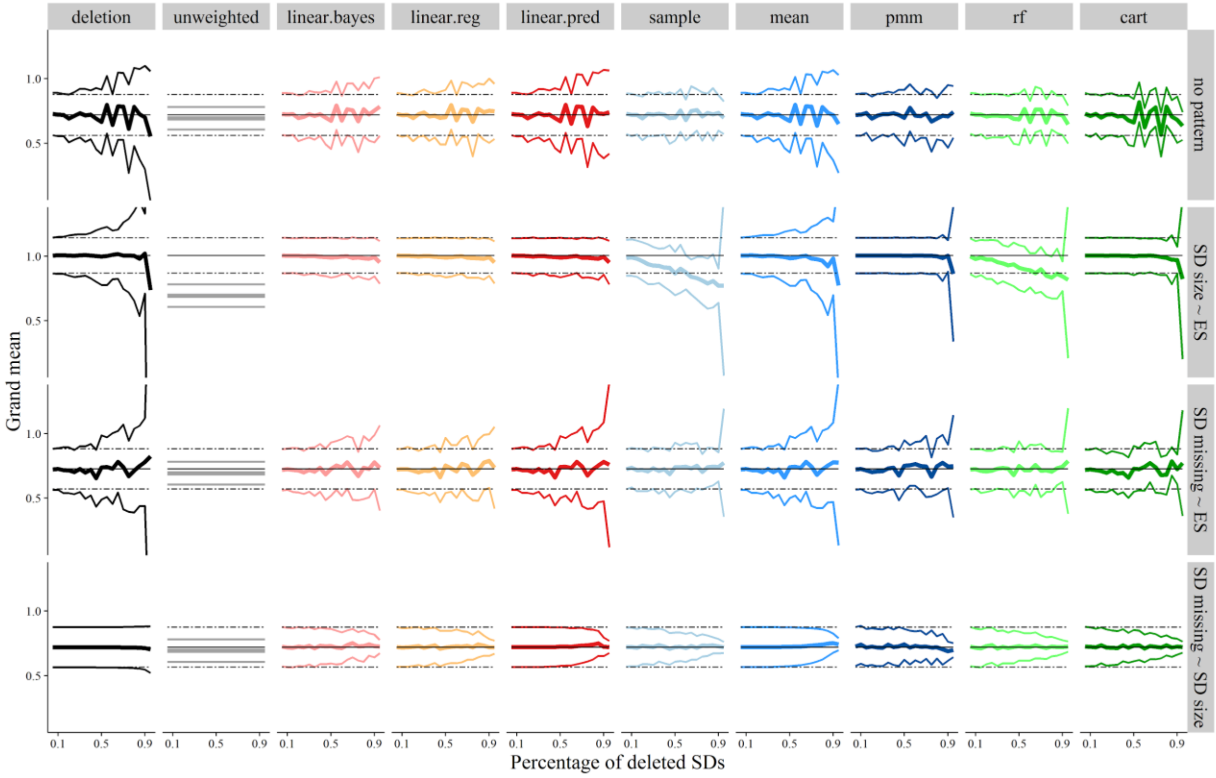


Figure | **7.4.** Grand mean estimates obtained after multiple imputation of missing standard deviations (SDs). Rows correspond to the different dataset characteristics: SDs were distributed and deleted at random ("no pattern"), SDs negatively correlated with the corresponding effect sizes (SD size ~ ES), the probability of deleting SDs negatively correlated with the corresponding effect sizes (SD missing ~ ES) or with the SD size (SD missing ~ SD size). In every artificial dataset (with 500 study cases), between 5% and 95% of the SDs in the treatment group were deleted (in steps of 5%) and then imputed with the methods described in Table 7.2. Black solid and dotted lines show the grand mean and confidence interval of the full dataset (i.e. with all SDs available). Coloured lines show the estimated grand means and confidence intervals obtained from the imputed data sets.

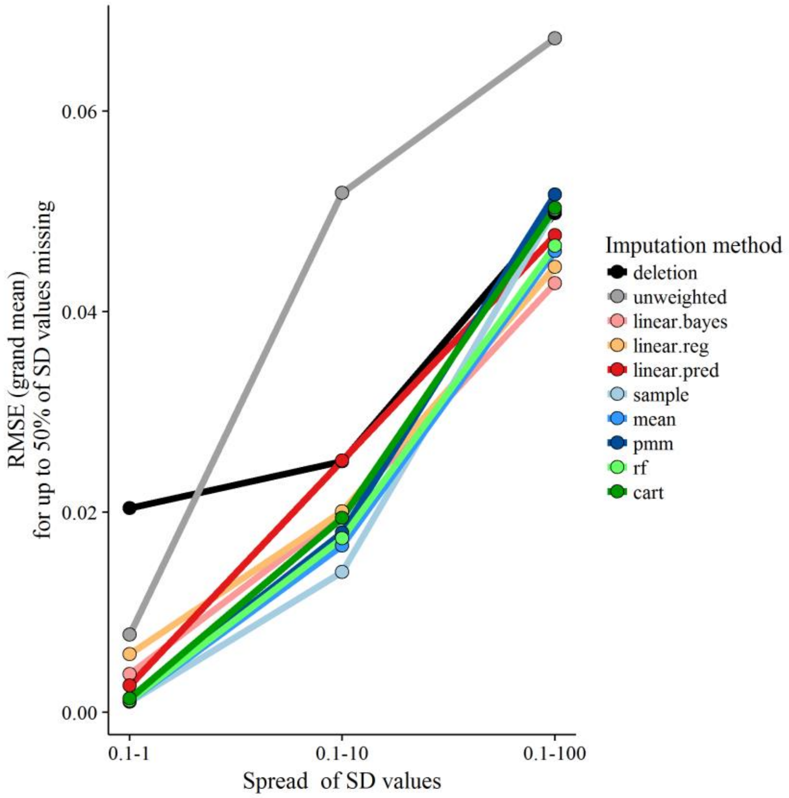
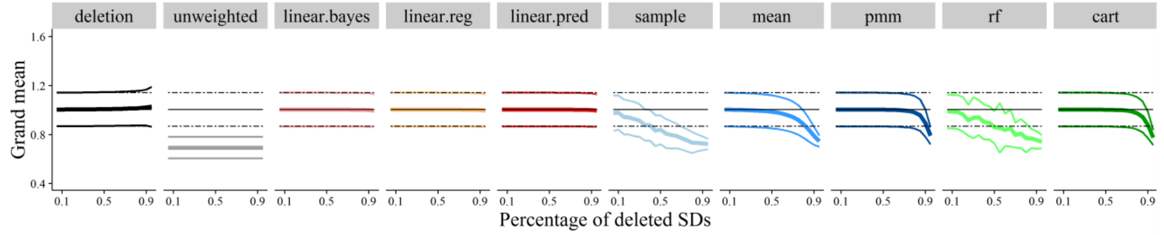


Figure | 7.5. Root mean square error (RMSE) between the grand mean of the full dataset and datasets with 5% up to 50% of imputed standard deviations (SDs). Applied imputation methods are described in Table 7.2. SDs were distributed and deleted at random and evenly distributed across the whole spread (shown on the x-axis). RMSE values were averaged across datasets of size 10, 20, 50, 100, 200 and 500 studies.

Imputation of missing SDs (SD ~ ES and missingness ~ SD size)



Imputation of missing sample sizes (sample size ~ ES and missingness ~ sample size)

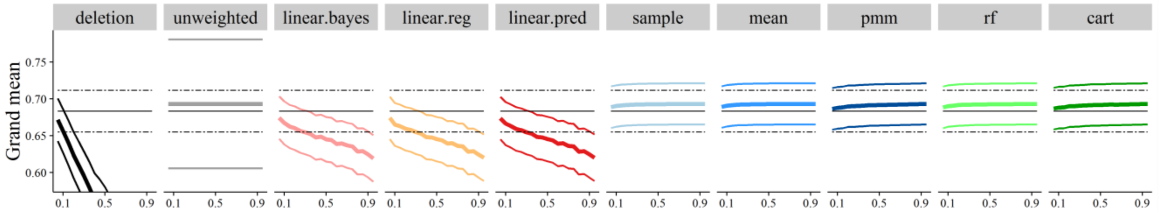


Figure | **7.6.** Grand mean estimates obtained after multiple imputation of missing standard deviations (SDs) and missing sample sizes when SDs or sample sizes correlated with the effect size and their chance to be deleted correlated with the SD or sample size. Effect size variances were calculated using eqn 7.2 for missing SDs and eqn 7.3 for missing sample sizes. In every artificial dataset (with 500 study cases), between 5% and 95% of the SDs in the treatment group were deleted (in steps of 5%) and then imputed with the methods described in Table 7.2. Black solid and dotted lines show the grand mean and confidence interval of the full dataset (i.e. with all SDs or sample sizes available). Coloured lines show the grand mean estimates and confidence intervals obtained from the imputed datasets.

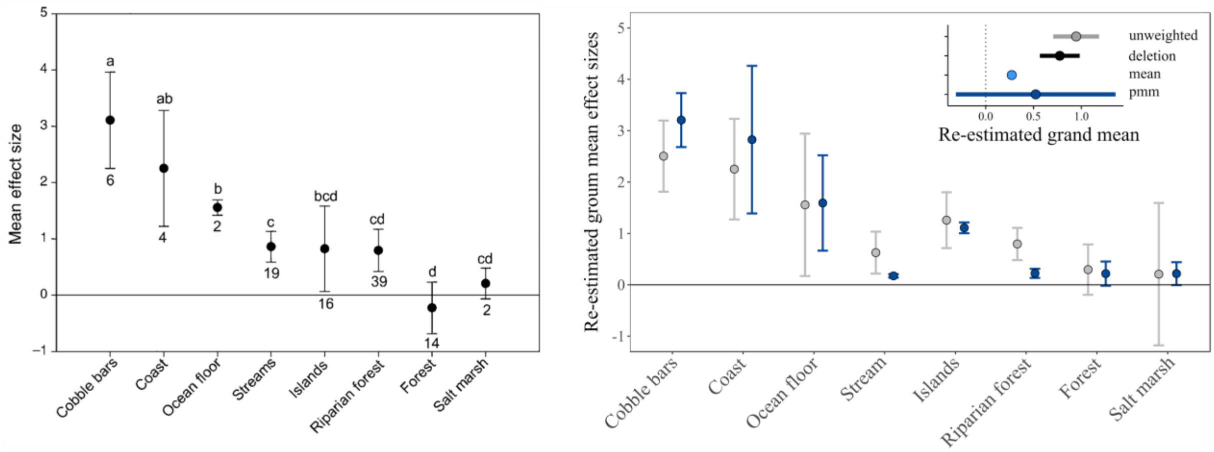


Figure | **7.7.** Re-analysis of the meta-analysis dataset provided by Marczak, Thompson & Richardson (2007) that tested the response of consumer biomass and density to resource subsidies in different habitat types. Left figure: because of missing standard deviations (SDs) the authors applied an unweighted analysis to determine mean effect sizes for the different habitat types (original Figure 1 in Marczak, Thompson & Richardson 2007). Large right figure: Re-estimated group mean effect sizes of the different habitat types using two imputation methods: unweighted meta-analysis (method *unweighted*) and a weighted analysis with missing SDs imputed via predictive mean matching (method *pmm*). Small figure in the top right corner: Re-estimated grand mean of this dataset using different imputation methods for the missing SDs: unweighted meta-analysis (method *unweighted*), weighted analysis while omitting incompletely reported studies (method *deletion*), weighted analysis with missing SDs imputed as the mean SD (method *mean*) and missing SDs imputed via predictive mean matching (method *pmm*).

8 | Synthesis

In the first and second study of this thesis I used meta-analytical tools across independent experiments, observations and the published literature to post-hoc analyse the effects of differences in the study design on the observed forest BEF relationships.

In study 1, I provided the most compelling evidence so far that the effects of species mixing on tree growth are inconsistent between experimental and observational research approaches across Europe (i.e. between national forest inventories, comparable plots in established forests and tree diversity experiments). In accordance with the findings of Duffy *et al.* (2017), the data obtained with observational approaches showed a generally positive biodiversity effect whereas the data from tree diversity experiments suggested that this effect is not (yet) significant.

In study 2, I summarized the published literature and found the effects of tree diversity on insect herbivory to depend on the studied aspect of herbivory and the mean annual temperature at the study site. To my knowledge, I thereby provided the first evidence that associational effects of tree species against pests could change along a gradient from boreal to tropical studies.

Unfortunately, the results of study 1 and 2 cannot directly be linked, because they were obtained from different data sets (i.e. from harmonized data sets and published studies), they investigated different measures of tree diversity (i.e. species mixing and Shannon diversity) and they differed in the covered geographical, environmental and compositional gradient. Thus, my results do allow the draw the generalization that tree diversity has positive effect on tree growth but does not affect insect herbivores.

If one wishes to use meta-analysis to test simultaneous relationships (i.e. covariation), for instance the effects of tree diversity on growth and herbivory, this analysis must be based on publication or data sets that reported all relationships of interest. This can severely restrict the meta-analysis data set, up to the point that similar, albeit slightly different relationships must be included in the dataset, which introduced more heterogeneity that must later be accounted for. Network meta-analyses can be applied to indirectly compare clinical treatments that were conducted in different studies if the control groups of those studies can be assumed to be somewhat similar (Li *et al.* 2011, Tonin *et al.* 2017). Whether this approach can be used to study multiple effects across different ecological publication has, to my knowledge, not yet be demonstrated. More complex relationships (that might be formulated as correlation or covariation matrices, that could be compiled from different publication) might be investigated with structural equation model-based meta-analyses (Cheung 2014, 2015). I am, again, not aware that this approach has yet been applied to ecological questions, although it might conceivably be useful to synthesize the multiple and multi-trophic effects of tree diversity on forest functioning.

In a recent review, Grossman *et al.* (2018) noted that, since the positive effects of biodiversity on ecosystem functioning are largely confirmed, research focus is now shifting towards the mechanisms that underlie the context-dependency of those BEF relationships. The meta-analysis of BEF relationships and their potential sources of heterogeneity in forest ecosystems is quite a developing field of research, because only nowadays there exists a sufficiently large data basis for synthesis analyses. Publication-based meta-analyses (*sensu stricto*), on the one hand, can only be done when there already exists a body of previously published studies on the focal BEF relationship. Standardized data sets that can be analysed with meta-analytical techniques, on the other hand, can often only be acquired in transnational research networks and must be collated according to reference protocols, as applied in for instance the grassland nutrient network (www.nutnet.umn.edu), the grassland and forest protocols of the climate action reserve (www.climateactionreserve.org/how/protocols), the FunDivEurope project (www.fundiveurope.eu), the global network of tree diversity experiments (www.treedivnet.ugent.be), the Center for Tropical Forest Science - Forest Global Earth Observatory (CTFS-ForestGEO, www.forestgeo.si.edu) and recently promoted by Trogisch *et al.* (2017).

All research and every policy or management recommendations should be based on all available evidence (Díaz *et al.* 2015). Also, future syntheses across complex, multi-trophic relationships (like in the case of context-dependent forest BEF relationships) can be only conducted on an extensive, complete and unbiased data basis. Therefore, published studies and data sets should, desirably, routinely report all measurements and relationships, regardless of their significance (as demanded in, for instance, Gerstner *et al.* 2017), together with information on the variability in the observed measurements and relationships, because this information can be used to weight individual studies in any subsequent meta-analysis.

In study 3, I clearly demonstrated that the (frequently applied) practice of omitting those studies that did not report effect sizes variances or sample sizes likely leads to biased grand mean effects and thus erroneous meta-analysis results. The results of my simulations showed that the imputation of those missing values can yield reliable grand mean estimates; under the consideration the supposed correlation structure of the meta-analysis data set.

In the first and second study of this dissertation, I did not encounter missing values, because these analyses were based on a harmonized raw dataset and a meta-analysis dataset obtained from only completely reported publications. The problem of missing variance estimates still is a prevailing problem in ecological meta-analyses. Exemplarily, McCrackin *et al.* (2017) found only 98 (9%) of the compiled 1.093 effect sizes, addressing the effects of reducing anthropogenic nutrient input in aquatic ecosystems, to be reported with the corresponding standard deviation. In this example, the authors thus applied an unweighted analysis on the full and a weighted analysis only on the subset of completely reported effect sizes. For this illustrative case, the third study of my dissertation actually indicates that imputing those missing values can still yield reliable and interpretable grand means, even if the proportion of reported standard deviations is seemingly small. I hope that these findings will, in the long term, lead to a more frequent application of imputation techniques and thereby to an increase in the quality and reliability of future meta-analyses; also outside of the field of ecology.

If ecological sciences should indeed suffer from a severe lack of reproducibility (as suggested by Schnitzer & Carson 2006), the imputation and re-analysis of published data could be one tool to evaluate the reproducibility (or generality) of ecological relationships and explore the sources of heterogeneity in study results (Schnitzer & Carson 2016, Fidler *et al.* 2017). Up to now, these meta-analyses in forest BEF research were restricted to only few univariate relationships but the ongoing accumulation of published forest BEF relationships and new statistical methods will likely increase the frequency and importance of global reviews (e.g. Grossmann *et al.* 2018), global analyses (e.g. Liang *et al.* 2016, Crouzeilles *et al.* 2016) and synthesis across multi-trophic levels of forest BEF relationships.

9 | Appendix

9.1 | Supplementary Materials

S4.1 | Text	Commentary publication advocating a raise in publication standards to foster the inclusion of primary studies into future meta-analyses (Gerstner *et al.* 2016).
S5.1 | Tables	Characteristics of the datasets included in the inventory, experimental and exploratory approach.
S5.2 | Table	List of all tree species that occurred in the inventory, experimental or exploratory data sets.
S5.3 | Figure	Histograms of the climate conditions that were covered by the inventory, experimental and exploratory research approach.
S5.4 | Table	Species composition that were shared between the three different research approaches.
S5.5 | Text	Sampling design of national forest inventory plots.
S5.6 | Figure	Locations of the mixed and monospecific plots in the inventory approach.
S5.7 | Text	Tree growth measurement in the exploratory platform.
S5.8| Table	Data preparation to compare the effects of species mixing across the three research approaches.
S5.9 | Figure	Histogram of Euclidean distances between assigned mixed and monospecific plots in the inventory approach.

S5.10 | Figure	Comparison of tree species responses to species mixing between the three research approaches when the data sets were restricted to only those community compositions and forest types that were shared between the compared research approaches.
S6.1 | Table	PRISMA checklist.
S6.2 | Text	Search protocol.
S6.3 | Figures	Funnel plot diagrams.
S6.4 | Data	Coding table, stored on the supplementary CD.
S7.1 | Data	List of reviewed meta-analyses.
S7.2 | Text	Description of the review of published meta-analyses.
S7.3 | Figure	The effects of ten methods to impute missing SDs on the grand mean under consideration of different dataset sizes.
S7.4 | Figure	The effects of ten methods to impute missing sample sizes on the grand mean under consideration of different dataset characteristics.
S7.5 | Text	Description of the imputation and re-analysis of a published meta-analysis dataset.

Methods in Ecology and Evolution

BRITISH ECOLOGICAL SOCIETY

Methods in Ecology and Evolution 2017, **8**, 777–784 doi: 10.1111/2041-210X.12758

COMMENTARY

Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting

Katharina Gerstner*[1,2], David Moreno-Mateos[3,4], Jessica Gurevitch[5], Michael Beckmann[2], Stephan Kambach[6,7], Holly P. Jones[8] and Ralf Seppelt[2,9]

[1]iDiv – German Centre for Integrative Biodiversity Research, 04103 Leipzig, Germany; [2]Department Computational Landscape Ecology, UFZ – Helmholtz Centre for Environmental Research, 04318 Leipzig, Germany; [3]Basque Centre for Climate Change – BC3, 48940 Leioa, Basque Country, Spain; [4]IKERBASQUE, Basque Foundation for Science, 48008 Bilbao, Basque Country, Spain; [5]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA; [6]Institute for Biology, Martin-Luther-University Halle-Wittenberg, 06099 Halle (Saale), Germany; [7]Department Community Ecology, UFZ – Helmholtz Centre for Environmental Research, 06120 Halle (Saale), Germany; [8]Department of Biological Sciences and Institute for the Study of the Environment, Sustainability, and Energy, Northern Illinois University, DeKalb, IL 60115, USA; and [9]Institute of Geoscience & Geography, Martin-Luther-University Halle-Wittenberg, 06099 Halle (Saale), Germany

---

Summary

1. Ecological and evolutionary research increasingly uses quantitative synthesis of primary research studies (meta-analysis) for answering fundamental questions, informing environmental policy and summarizing results for decision makers.

2. Knowing how meta-analysis works is important for researchers so that their research can have broader impact. Meta-analytic thinking encourages scientists to see single primary research studies as substantial contributions to a larger picture.

3. To facilitate inclusion in a meta-analysis, relevant primary research studies must be found and basic information about the methods and results must be thoroughly, clearly and transparently reported. While many published papers provide this information, it is common for essential data to be omitted, leading to study exclusion from meta-analyses.

4. We provide guidelines for correctly reporting basic data needed from primary studies in ecology and evolutionary biology so that they can be included in meta-analyses, together with examples that show how data should be reported to enable calculation and analysis of effect sizes, and how data should be made accessible.

5. These guidelines are important for reporting research results in general, whether or not results are included in subsequent meta-analyses, because they are necessary for the interpretation and assessment of study outcomes. Increased implementation of these guidelines by authors, editors and publishers, and reinforcement by funders, will foster higher quality and more inclusive syntheses, further the goals of transparency and reproducibility in science, and improve the quality and value of primary research studies.

Key-words: data reporting, meta-analysis, open science, quantitative review, reproducibility, research synthesis, transparency

Introduction

Researchers in ecology and related disciplines increasingly use quantitative synthesis of primary research studies for answering fundamental questions, testing hypotheses, informing environmental policy and summarizing results for decision makers (Cadotte, Mehrkens & Menge 2012; Pullin 2012; Mengersen, MacNeil & Caley 2015). To accomplish these goals, there are formal guidelines and statistical methods using meta-analysis to summarize results of independent studies and analyse general trends, as well as for evaluating factors that may cause heterogeneity in outcomes among studies (Borenstein *et al.* 2009; Koricheva, Gurevitch & Mengersen 2013).

As meta-analyses have become increasingly common and important in ecology and evolutionary biology, it is useful for all researchers to understand how a meta-analysis works. This knowledge not only enables scientists to interpret and evaluate published meta-analyses, but also allows them to make their own research accessible for meta-analysts so that their results can be incorporated and interpreted in the broader context of research on the questions they are addressing. Meta-analytical

*Correspondence author. E-mail: katharina.gerstner@idiv.de

thinking allows scientists to see single primary research papers as essential contributions to a larger picture within a research topic (Schmidt 1992; Nakagawa & Cuthill 2007). Using meta-analysis, researchers can consider how the results of an individual study relate to or contrast with others from different geographical areas or ecological systems, how new results complement or contradict earlier findings, and how they will be seen in the context of future findings. Thus, meta-analytical thinking increases the intellectual impact of individual studies and makes them more long-lasting. This is particularly compelling considering that about 27% of published individual research studies in natural sciences and engineering are never cited (Larivière, Gingras & Archambault 2009).

Here, we suggest specific guidelines for future primary research studies in ecology and evolutionary biology, amplifying and highlighting previous calls for higher publication standards including improved protocols for reporting procedures and outcomes of ecological studies (Hillebrand & Gurevitch 2013; Zuur & Ieno 2016). We also provide examples of how certain types of data should be reported to allow calculation of effect sizes from them, which is essential for meta-analyses. By implementing these guidelines for reporting research results, higher quality syntheses and meta-analyses of the current state of knowledge will be enabled. In this way, future meta-analyses will contribute to developing and answering new research questions in the domains of ecology, evolutionary biology, conservation biology and environmental science.

The principles of meta-analyses

In order to be jointly analysed in a meta-analysis, the outcome of each study must be expressed on a common scale. This measure of outcome, called 'effect size', includes information on the direction and magnitude of an effect of interest from each study. 'Effect size thinking' has been encouraged by various authors even in reporting primary research study results because it emphasizes outcome magnitude and direction, in contrast to *P*-values, which indicate neither (Arnqvist & Wooster 1995; Nakagawa & Cuthill 2007). The sampling variance of the effect size expresses the precision with which the effect is estimated. These effect size measures can then be combined across studies, taking into account the precision with which each is estimated. One can then estimate an overall mean effect and confidence intervals around that mean effect, test whether the overall effect differs significantly from zero, assess whether the outcomes of the studies are heterogeneous, and if so, test hypothesized categorical or continuous covariates (if any) to account for that heterogeneity.

Originally, meta-analysis was developed in the social sciences and medicine in the late 1970s and first introduced into ecology in the early 1990s (Koricheva, Gurevitch & Mengersen 2013). Meta-analyses in ecology and evolutionary biology not only share many aspects with those in medicine and the social sciences, but also differ substantially, for example in the types of research questions addressed and the data structures typically encountered in primary studies. A fundamental difference between meta-analyses in medical and ecological research is that the former are generally focused on specific estimates of the efficacy of a drug or surgical interventions, while the latter are often concerned with summarizing much larger and varied groups of studies to understand factors associated with the heterogeneity of their outcomes. Ecological meta-analyses, thus, are largely concerned with statistical issues involved with accounting for heterogeneity among studies rather than with estimating a single mean effect across all studies. For example, meta-analyses in ecology and evolution often encounter non-independence among studies due to shared phylogeny (Nakagawa & Santos 2012). Moreover, ecological and environmental primary research is characterized by a diversity of model systems, geographical variability and stochasticity inherent to data collected under less controllable conditions, for example in the field. Thus, ecological data are likely to vary according to factors like climate, geomorphological processes, soil conditions and local or global human activities that might also affect the data used in meta-analyses. Meta-analyses can be a powerful tool in assessing the impact of such factors when results are combined across studies but synthesis can be hampered by lack of transparency in reporting results, unstandardized or missing data descriptions, and missing or inadequate metadata describing data characteristics and collection protocols (Hillebrand & Gurevitch 2013).

Meta-analysis requires a major search and data extraction effort from primary research studies, and crucially depends on primary research studies to contain relevant keywords to enable studies to be located (Côté *et al.* 2013), and on the transparency and usability of reported data (Nakagawa & Cuthill 2007; Nakagawa & Parker 2015). Some previous recommendations to increase the quality of ecological studies, their analysis and the clarity with which they are reported, go back decades (Fowler 1990) while more recent efforts have urged study reproducibility and transparency (Goodman, Fanelli & Ioannidis 2016; Parker, Nakagawa & Gurevitch 2016b; Parker *et al.* 2016a). Numerous reviews in ecology and evolution find that as many as half of published articles lack key pieces of information regarding statistical relationships (Parker *et al.* 2016a). Moreover, important metadata is often not sufficiently reported (Hillebrand & Gurevitch 2013; Roche *et al.* 2015; Zuur & Ieno 2016) hampering the use of metadata as covariates in a meta-analyses.

If primary research papers are to be made accessible to meta-analyses, there is a need to promote 'meta-analysis thinking' which goes beyond 'effect-size thinking' (Nakagawa & Cuthill 2007) by considering how the paper will be found in a literature search, and how clearly and completely the procedures, analyses and results are reported, including relevant covariates that characterize the study or were used in the analyses.

Guidelines for meta-analytical thinking

GETTING FOUND IN A LITERATURE SEARCH

Meta-analyses are most often accompanied by a systematic literature search aiming at a representative sample of existing primary research papers. Identifying a search strategy is a first crucial step (Côté *et al.* 2013). The number of relevant papers may vary strongly depending on the data sources, such as standard search engines (e.g. Web of Science, SCOPUS, Google Scholar; Beckmann & von Wehrden 2012) or grey literature, and on the keywords used for the literature search (cf. fig. 4.1 in Côté *et al.* 2013). The number of studies found can be increased by using more keywords. However, just using a greater number of keywords also increases the number of 'hits' identified that are without actual relevance for the selected research question. Hence, for a meta-analyst, there is a trade-off between false positives (studies that are not relevant to the questions addressed by the synthesis) and false negatives (relevant studies that are missed by the search).

To increase the likelihood of primary research studies being identified in a literature search, authors should carefully consider the choice of title, abstract content and keywords to find a balance between being broad enough to be found through a keyword search and being specific enough to be identified as relevant. Abstract and title should not solely focus on one key finding, but briefly capture the full experiment, and the use of technical terms not recognized in other (sub)disciplines should be supplemented by generally recognized synonyms (cf. Table 1).

To improve the chances of a study being incorporated in a meta-analysis, open access papers are the gold standard. If a publication cannot be made open access, many journals allow users to host a manuscript in an unedited form (but permission depends on the policy and agreements established by the publisher). Online platforms, such as ResearchGate or Mendeley, substantially reduce the effort required for individual authors to provide full text access.

REPORTING USABLE OUTCOMES OF PRIMARY RESEARCH TO ENABLE CALCULATION OF EFFECT SIZES

Accurate data extraction from primary research studies is essential to calculate effect sizes and is one of the most time-consuming parts of a meta-analysis. The outcome and legitimacy of a meta-analysis depends on accurate and complete reporting of study outcomes. Effect sizes should summarize the results of each study on the same scale and in an unbiased manner. Many different effect size metrics can be used, from well-known to less familiar, each suitable for specific purposes and data conditions (Osenberg, Sarnelle & Cooper 1997; Nakagawa & Cuthill 2007; Mengersen & Gurevitch 2013). Data needed from each study to calculate effect sizes depend on the research question, the data structure and the specific metric of effect size chosen. The most commonly used effect sizes in ecology are standardized mean differences, response ratios and correlation coefficients (using Fisher's Z transformation). Response ratios and (standardized) mean differences are often used in ecological research syntheses when the goal is estimating the magnitude of the effect of an experimental treatment on a continuous response variable (such as biomass). This is the case where one wishes to compare means for experimental and control groups from studies where means and standard deviations are reported. These measures can also be used for non-experimental data where one is comparing means of two groups as a measure of 'effect'. Odds and risk ratios can be used for binary response variables (alive/dead, pollinated/not pollinated) but are much less common in ecology than in medical meta-analysis. If the relationship between two continuous variables is of interest, correlation coefficients are often appropriate; slopes from simple linear regressions are problematic as effect sizes for several reasons, including calculation on very different scales among studies (Rosenberg, Rothstein & Gurevitch 2013). It is necessary to understand the sampling distribution of the effect size metric used so that one can correctly calculate a measure of sampling variance to be used in subsequent analyses, including weighting the outcomes according to their precision (Gurevitch, Curtis & Jones 2001).

Papers reporting on research outcomes from primary research should generally include basic information on means, sample sizes and measures of variation to be useful in a meta-analysis, or if correlation coefficients are reported or calculated, the sample sizes should be provided (Fig. 1). Ideally, raw data are provided. If raw data are presented in figures, points should be clearly distinguishable and non-overlapping. Study design should be clearly documented, including hierarchical designs and any aggregation of (raw) data (e.g. do the means and confidence intervals correspond to individual organisms or are they plot means?). It is essential to identify the measure of variation reported (e.g. standard error, standard deviation or 95% confidence interval) in the text, as well as in figures and table captions.

In order to understand the outcome of a study, it is valuable to report or be able to calculate effect sizes and their variation. This provides readers with more complete information not only of statistical significance, but also with the magnitude and variation in effect sizes (Nakagawa & Cuthill 2007). Furthermore, it is important to report all of the results, including the statistically non-significant ones. 'Not significant' does not mean 'non-informative'. Reporting only significant results ('*p*-hacking') create biases ('publication bias'), and those biases are magnified when results are combined across studies (Rothstein, Sutton & Borenstein 2006). It is a widespread misconception that reporting on significant findings may increase the probability of getting a paper accepted (Koricheva 2003), and non-significant results can be of great relevance in a broader context (Parker *et al.* 2016a). We thus support and amplify the Center for Open Science Transparency and Openness (TOPS) guidelines ('OSF|Tools for Transparency in Ecology and Evolution (TTEE)', 'COS|Openness, Integrity, and Reproduciblity') and the checklist published by Hillebrand & Gurevitch (2013) to include reports on all results, regardless of statistical significance and direction of the effect.

A common goal of ecological meta-analyses is the analysis of effect sizes and causes of variation in study outcomes. Therefore, another crucial step alongside the data extraction for calculation of effect sizes is the data extraction of relevant covariates (moderators). Variation in study outcomes may be

Table 1. : Core issues, action-items and reasoning to be considered in primary research studies in ecology and evolution to be beneficial for use in meta-analysis studies.

Common problems found in primary research papers	What to do?	Why is this important?
Getting found in a literature search		
Study is not indexed in major bibliographic databases (while master or doctoral theses are indexed, "grey" literature, e.g. NGO reports, are difficult to find).	Journals such as Research Ideas and Outcomes (RIO) promote publishing any part of the research cycle (project proposals, data, methods, workflows, software, project reports) and consequently provide access to "grey" literature, too. Alternatively, meta-analysts may use less restrictive search engines to Web of Science, like Google Scholar or Scopus.	Studies that are not indexed in major bibliographic databases will usually not be found or considerable effort is needed.
Studies might be overlooked if title, abstract and keywords are too generic.	Consider your abstract and title to be distinctive from a review or opinion paper in the same field. In the title and abstract, use words that describe the main finding of the study, the geographical context, the methodology, and if possible the main covariates used in the analysis	Relevant studies might not be identified and do not reveal that codable information is presented.
Studies might be overlooked if title and abstract are too specific, i.e. highlight major findings and omit additional data and tests.	Consider possible broader questions, the findings might contribute to (without overselling). Use this for title or a final synthesis-section or -sentence in the abstract. Be as brief but comprehensive as possible by itemizing tests, hypotheses, or data in the abstract, including non-significant results.	Relevant studies might not be identified, as the presented findings might be seen as out of scope.
A paper cannot be accessed by the meta-analyst's research institution.	If a publication cannot be made open-access, some journals allow users to host a manuscript in an unedited form. Otherwise, respond timely to pdf requests.	The study is identified as relevant, but the results cannot be used, and the paper will be listed as "not accessible" in the meta-analysis study.
Reporting usable outcomes to enable calculation of effect sizes		
Deficiencies in reporting statistics	Present either raw data or summary statistics for the response, including means, variation around the mean and number of samples for both the control and treatment groups. If raw data are presented in figures avoid overlapping points in scatterplots, use transparency and indicate full overlap. Aggregation of (raw) data should be clearly documented (e.g. are the means and confidence intervals the means of individual organisms or plot means). In regression-type analyses either raw data or regression slopes should be presented along with their confidence intervals, in addition to the coefficient of determination R^2.	Missing statistic information hampers effect size calculation.
Units are not reported.	Exhaustively report what was measured and how it was measured in figures, tables, and text; always report what the measure of variation is (SE, 95% confidence intervals, or standard deviation, based on what sample size, etc.).	Missing information may hamper the calculation of effect sizes and its variance.
Negative or non-significant results are not reported or lack sufficient detail.	Report all results, regardless of statistical significance and direction of the effect. If results are considered as not being of direct interest for the research question, results can be put into the supplementary material.	Reporting negative or non-significant results prevents publication bias and enables replicability of the analyses.
Deficiencies in describing the study and experimental design.	Include the number and sizes of sampling sites, plots within sampling sites, and replicates within plots. In experimental studies, include the nature of replication; full details of how hierarchical designs were accounted for in the analyses (e.g. blocking, split plot designs, nested designs).	The number of sampling sites and plots may be used in effect size calculation and is useful to understand the primary research study, to make (spatial) comparisons across climates and regions, and facilitates using geographical information in larger-scale syntheses.
Study design is reported in previous papers by same authors, worst case in an inaccessible technical paper.	Be as specific as possible in describing the study design, use supplementary material section to report details even if you repeat information given in previous papers.	Study results cannot be related to the original plots/observations/data and the corresponding moderators in the meta-analysis.

Table 1. (continued)

Common problems found in primary research papers	What to do?	Why is this important?
Missing exact geographic location.	Include coordinates, geographic system, altitude, and depth as accurately as possible.	Studies cannot be mapped nor related to regional climate data, general vegetation and other mapped data. Study locations may also be needed to check whether the selection of primary research studies are geographically biased or representative with respect to environmental and socio-economic context.
Full environmental features of the study area are not reported.	Include information on the climate, soil, elevation, and the exact type of ecosystem (there are many kinds of wetlands or forests) and other study-specific and system-specific results.	Complete reporting of such metadata helps with the interpretation of the results of individual studies. Conditions could be used as a covariate in the meta-analysis or help to group studies.
Historical conditions unknown.	If your paper relates to environmental degradation or in general refers to any kind of change processes, describe in detail how the study area or system was affected and when the degradation occurred (e.g., duration, start and end dates,).	Complete reporting of such metadata helps with the interpretation of the results of individual studies. Conditions could be used as a covariate in the meta-analysis or help to group studies.
No reference system for species taxonomy/ species names given.	Adapt the species names to an accepted taxonomic list (e.g. the plant list) and state which taxonomy at which time point was adapted.	Reference systems are needed to correctly match datasets on species occurrences and abundances from various primary research studies.
Spatial scale is not sufficiently or correctly reported.	Include information about spatial grain (e.g., size of plots), focus and extent. Include the size of the whole experimental area, if exists, and the study area.	Complete reporting of spatial scale facilitates the analysis of scale-dependence of results, allows to estimate a "study area" and add additional georeferenced data which might explain observed patterns and explain the geographical reach of the meta-analysis.
Temporal scale is not sufficiently reported.	Include information about the exact start and end of observations or experiments.	Temporal information might be needed to analyse whether observed patterns are consistent over time to estimate whether community dynamics are responsible for the observed patterns, and to distinguish short-term from long-term results.
Increasing data accessibility		
Databases are not easily accessible.	Provide data in a digital format in data sheets (e.g., xls, csv, shp) rather than text based formats (pdf).	Increased effort is needed to gain access to the data, which may lead to ambiguity and error in subsequent analysis steps.
Data are no longer available or link to data repository does not work.	Ensure that the data are still available in five, 20, or 50 years using public, global repositories that guarantee long-time maintenance, at best keeping track on improved versions and updates. Examples are datadryad.org, figshare.com, pangea.de which also provide digital object identifier (DOI) for data set(s).	Filing data at the authors', publishers' or institutions' websites may not guarantee longer-term availability as they are updated or become unsupported, fail or are taken offline.
Meta-data are not provided or are incomplete.	Ensure that metadata are completely provided, e.g., principal investigator, responsible institution, site location including geographic coordinates, plot size, scale of experiment, year, and environmental conditions. Available meta-data standards (i.e. Dublin Core) can provide guidelines.	Alongside results, various aspects of a study are of relevance that further help to use and analyse the study results. A study might even drop out from a meta-analytic analysis if important information is missing, which cannot be imputed.
Data requests are not responded to or responses are delayed.	Respond to data requests as soon as possible. Researchers have an obligation to their funding and salary sources to report their results fully and accurately, and to the organisms and systems they study to make their results available to science.	Timely responses to data requests allow having meta-analytical databases of higher quality, with substantially more information than with few or delayed responses.

due to biologically meaningful and important covariates (organism traits, climate, population density, etc.) or to methodological covariates (study duration, experimental conditions, source of material used). Potential confounding of covariates must be directly discernible (e.g. all woody species were grown in the ground, all herbaceous species in pots, so one cannot separate growth form from growth conditions). Detailed description of experimental methods and study design should include the number and sizes of sampling sites, plots within sampling sites, replicates within plots and study duration. Description of study area should include exact geographical location (as geographical coordinates or as a map) as it may be needed for various reasons: First, to access regional climate data, general vegetation and other mapped data. Second, to map study locations and check whether the selection of primary research studies are geographically biased (Martin, Blossey & Ellis 2012) or representative with respect to environmental and socio-economic context (Margulies *et al.* 2016). Further information should be provided on the environmental context such as climate, soil, elevation, and the specific type of ecosystem and other study-specific and system-specific results. Other contextual features, such as land-use history, might be useful to interpret results and make them comparable with other sites or studies. If a study investigates effects of environmental change, describing how the study area or system was affected (duration of the impact and start and end date, any measure of the intensity of the alteration) helps with the interpretation of the results of individual studies and may be useful as a covariate in meta-analysis or help to group studies. To correctly match datasets on species occurrences and abundances from primary research studies authors must report species names according to a documented taxonomic authority (e.g. the plant list) and state which taxonomy at which time point was adapted.

In ecology, the spatial scale of observation deserves specific treatment (Whittaker 2010; Chase & Knight 2013). However, primary research studies frequently fail to report the details of the spatial scale of the study by incorrectly distinguishing between spatial scale components grain, focus, and extent, and even erroneously reporting the spatial scale to which the summary statistics belong (Whittaker, Willis & Field 2001). Grain denotes the size of the analytical unit. Focus is the area or inference space represented by each data point and thus represents the scale at which the grains are aggregated or the scale at which a mean is calculated. Extent is the scale at which the entire set of sample-units is analysed (Scheiner *et al.* 2000). While grain is frequently reported, focus and extent are often missing or descriptions do not allow clear distinctions among the spatial scale components. Such ambiguity hampers the analysis of spatial scale effects, and makes among-study comparisons less reliable.

INCREASING DATA ACCESSIBILITY

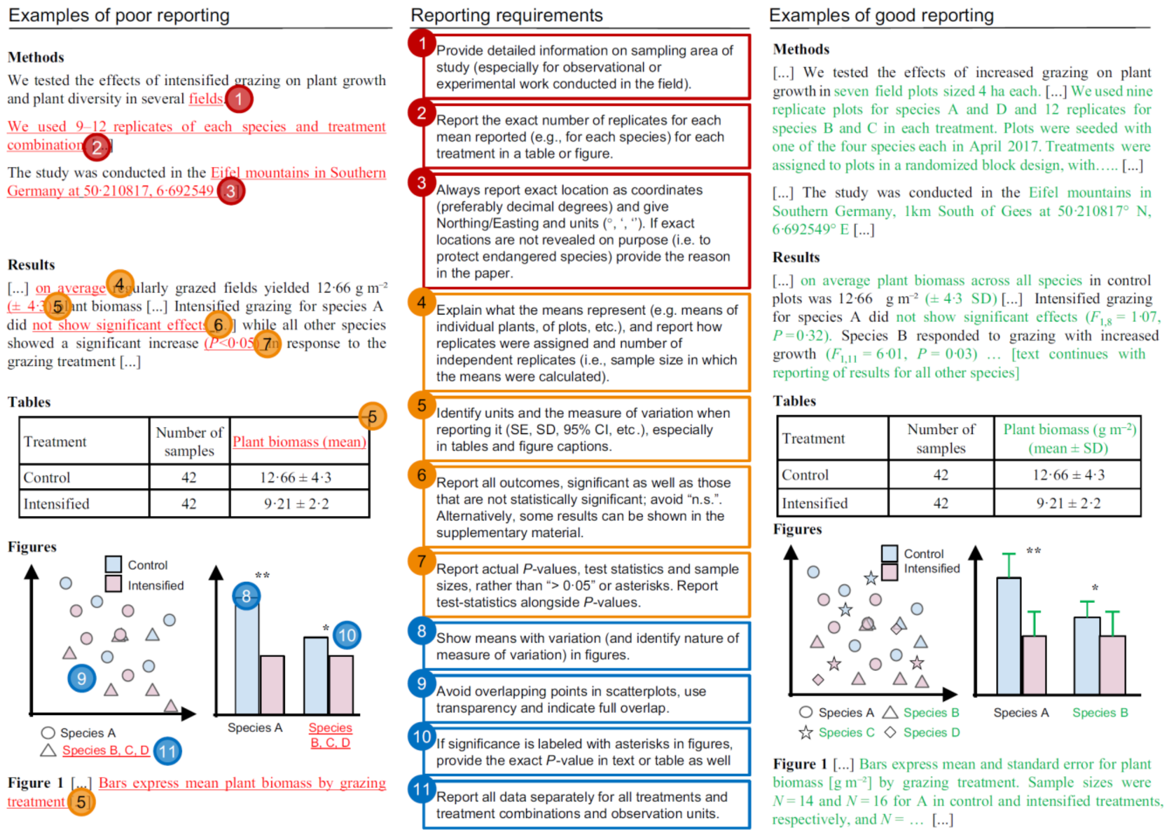
Complete access to all data in published studies has been a much-debated issue (e.g. Gewin 2016). For meta-analyses, however, it is extraordinarily useful. Although most results in primary research studies are provided in tables or figures and can be extracted by the meta-analysts, the underlying data should be made available online, because they may reduce effort, ambiguity and error in subsequent analysis steps. Access to archived data is guaranteed if digital object identifiers are provided and data stored in global repositories, like figshare (https://figshare.com), Dryad (https://www.datadryad.org) or Pangaea (https://www.pangaea.de). In contrast, filing data at the authors', publishers' or institutions' websites may not guarantee longer term availability as they are updated or become unsupported, fail or are taken offline. It is critical to include sufficient metadata along with the archived data. The paper on the publisher's homepage should be always linked with this repository. This 'Open Science' approach is now widely supported and is requested by many journals and funding agencies. However, it still lacks full support in the academic reward system (Nosek *et al.* 2015).

To obtain the required data missing from published papers, the last option for a meta-analyst is to request the data directly from the study authors. This is usually the most time consuming strategy as responses are often delayed, ignored or data access is only granted in exchange for co-authorship. However, researchers have an obligation to their funding and salary sources to report their results fully and accurately, and to the organisms and systems they study to make their results available to science. As data reporting standards in primary research studies improve in the future, the need for data requests will likely diminish.

Discussion

Comprehensive reporting of results is important for understanding the primary research study and also facilitates future syntheses, as it will provide a more general understanding across different systems or experimental replicates. This may be seen as an opportunity for primary research authors as it will increase study's transparency and clarity, and increase the options to use a study's data and findings, thereby increasing citations and broaden its impact. It could also stimulate the willingness of scientists to conduct more meta-analyses once selecting and extracting the data seem less daunting. We estimate that the effort required to address these recommendations would not impose a major burden to the authors compared to the effort required to collect, analyse and publish the data.

There are also implications for editors, publishers and funders. To make the value of primary research longer lasting and have broader impact, journal editors, reviewers and publishers should provide published guidelines that require the listed information on the results of the study be included upon initial submission. For example, studies should report the size of experiment, constraints, environment and scale of investigation (some journals like 'Nature' and 'Nature Communications' already have these basic standards). Further, journals should require that data supporting the results in papers published in its journals will be archived in an appropriate public archive (e.g. '*Methods*



Examples of poor reporting

Methods

We tested the effects of intensified grazing on plant growth and plant diversity in several fields ①

We used 9–12 replicates of each species and treatment combination ② [...]

The study was conducted in the Eifel mountains in Southern Germany at 50·210817, 6·692549 ③ [...]

Results

[...] on average ④ regularly grazed fields yielded 12·66 g m^{-2} (± 4·3) ⑤ plant biomass [...] Intensified grazing for species A did not show significant effects ⑥ [...] while all other species showed a significant increase (*P*<0·05) ⑦ in response to the grazing treatment [...]

Tables

⑤

Treatment	Number of samples	Plant biomass (mean)
Control	42	12·66 ± 4·3
Intensified	42	9·21 ± 2·2

Figures

Figure 1 [...] Bars express mean plant biomass by grazing treatment ⑤ [...]

Reporting requirements

① Provide detailed information on sampling area of study (especially for observational or experimental work conducted in the field).

② Report the exact number of replicates for each mean reported (e.g., for each species) for each treatment in a table or figure.

③ Always report exact location as coordinates (preferably decimal degrees) and give Northing/Easting and units (°, ‘, ‘’). If exact locations are not revealed on purpose (i.e. to protect endangered species) provide the reason in the paper.

④ Explain what the means represent (e.g. means of individual plants, of plots, etc.), and report how replicates were assigned and number of independent replicates (i.e., sample size in which the means were calculated).

⑤ Identify units and the measure of variation when reporting it (SE, SD, 95% CI, etc.), especially in tables and figure captions.

⑥ Report all outcomes, significant as well as those that are not statistically significant; avoid “n.s.”. Alternatively, some results can be shown in the supplementary material.

⑦ Report actual *P*-values, test statistics and sample sizes, rather than “> 0·05” or asterisks. Report test-statistics alongside *P*-values.

⑧ Show means with variation (and identify nature of measure of variation) in figures.

⑨ Avoid overlapping points in scatterplots, use transparency and indicate full overlap.

⑩ If significance is labeled with asterisks in figures, provide the exact *P*-value in text or table as well

⑪ Report all data separately for all treatments and treatment combinations and observation units.

Examples of good reporting

Methods

[...] We tested the effects of increased grazing on plant growth in seven field plots sized 4 ha each. [...] We used nine replicate plots for species A and D and 12 replicates for species B and C in each treatment. Plots were seeded with one of the four species each in April 2017. Treatments were assigned to plots in a randomized block design, with….. [...]

[...] The study was conducted in the Eifel mountains in Southern Germany, 1km South of Gees at 50·210817° N, 6·692549° E [...]

Results

[...] on average plant biomass across all species in control plots was 12·66 g m^{-2} (± 4·3 SD) [...] Intensified grazing for species A did not show significant effects ($F_{1,8}$ = 1·07, P = 0·32). Species B responded to grazing with increased growth ($F_{1,11}$ = 6·01, P = 0·03) … [text continues with reporting of results for all other species]

Tables

Treatment	Number of samples	Plant biomass (g m^{-2}) (mean ± SD)
Control	42	12·66 ± 4·3
Intensified	42	9·21 ± 2·2

Figures

Figure 1 [...] Bars express mean and standard error for plant biomass [g m^{-2}] by grazing treatment. Sample sizes were N = 14 and N = 16 for A in control and intensified treatments, respectively, and N = … [...]

Fig. 1. Examples of often encountered, erroneous descriptions of methodology (red points), reported results in text, tables (orange points) and figures (blue points). Incorrect elements are highlighted as red. The centre column provides instructions for correctly described methods and reported results and the right column illustrated how the example text, table or figure should be prepared to be of optimum use in a meta-analysis.

in *Ecology and Evolution*’). Editors and publishers should consider treating every study used in a meta-analysis as a full citation by including a second list of references either within the main paper after the primary reference list in a reduced font size (e.g. ‘*Global Ecology and Biogeography*’) or in the online version of the paper (e.g. ‘*Nature*’) instead of listing them in the appendices or supplementary material. This will ensure that they are indexed and earn the deserved citation (McGill, Dornelas & Field 2016). This is, however, highly controversial in the scientific publication world. Space limitations should not be a prohibitive factor in an era in which printed journals have become a rarity and data syntheses are increasingly essential.

Data-based synthesis will foster ecological research and provide decision makers with valuable knowledge, institutionalized in synthesis centres in ecology, biology, biodiversity and other fields (see the Synthesis Consortium; http://synthesis-consortium.org). Other forms of meta-studies such as systematic reviews would also surely benefit from better result reporting as suggested by our guidelines (Magliocca *et al.* 2015). Scientists need to acknowledge that each piece of research could be more than just a publication on a list; it could be an important part of a larger picture. Finally, the more studies that can be included in meta-analyses, the more likely we will be able to glean a more comprehensive understanding of the generalities and tests of theory that meta-analyses in ecology often seek to build.

Authors’ contributions

K.G., R.S., M.B. and J.G. conceived the idea; K.G., D.M.M., J.G., M.B., S.K. and H.P.J. provided synthesised in-depth experience from conducting meta-analyses. KG and RS led the writing of the manuscript with the support of D.M.M., M.B., J.G., S.K. and H.P.J.

Acknowledgements

This work was supported by the National Socio-Environmental Synthesis Center (SESYNC) under funding received from the National Science Foundation DBI-1052875, by the Helmholtz Centre for Environmental Research, and by the Synthesis Centre (sDiv) of the German Centre for Integrative Biodiversity Research (DFG FZT 118). We acknowledge funding from the Helmholtz Association (Research School ESCALATE, VH-KO-613, M.B., S.K.), the Germany Federal Ministry of Education and Research (GLUES, 01LL0901A, K.G.) and the National Science Foundation (1119891, J.G.). We thank Eduardo S.A. Santos and one anonymous referee for their valuable comments on the manuscript.

Data accessibility

Manuscript does not include any data.

References

Arnqvist, G. & Wooster, D. (1995) Meta-analysis: synthesizing research findings in ecology and evolution. *Trends in Ecology & Evolution*, **10**, 236–240.

Beckmann, M. & von Wehrden, H. (2012) Where you search is what you get: literature mining–Google Scholar versus Web of Science using a dataset from a literature search in vegetation science. *Journal of Vegetation Science*, **23**, 1197–1199.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. & Rothstein, H.R. (2009) *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd, New York, NY, USA.

Cadotte, M.W., Mehrkens, L.R. & Menge, D.N.L. (2012) Gauging the impact of meta-analysis on ecology. *Evolutionary Ecology*, **26**, 1153–1167.

Chase, J.M. & Knight, T.M. (2013) Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecology Letters*, **16**, 17–26.

COS|Openness, Integrity, and Reproduciblity. Available at: https://cos.io (accessed 24 October 2016).

Côté, I.M., Curtis, P.S., Rothstein, H.R. & Stewart, G.B. (2013) Gathering data: searching literature and selection criteria. *Handbook of Meta-Analysis in Ecology and Evolution* (eds. J. Koricheva, J. Gurevitch & K. Mengersen), pp. 37–51. Princeton University Press, Princeton, NJ, USA.

Fowler, N. (1990) The 10 most common statistical errors. *Bulletin of the Ecological Society of America*, **71**, 161–164.

Gewin, V. (2016) Data sharing: an open mind on open data. *Nature*, **529**, 117–119.

Goodman, S.N., Fanelli, D. & Ioannidis, J.P. (2016) What does research reproducibility mean? *Science Translational Medicine*, **8**, 341ps12.

Gurevitch, J., Curtis, P.S. & Jones, M.H. (2001) Meta-analysis in ecology. *Advances in Ecological Research*, **32**, 199–247.

Hillebrand, H. & Gurevitch, J. (2013) Reporting standards in experimental studies. *Ecology Letters*, **16**, 1419–1420.

Koricheva, J. (2003) Non-significant results in ecology: a burden or a blessing in disguise? *Oikos*, **102**, 397–401.

Koricheva, J., Gurevitch, J. & Mengersen, K. (2013) *Handbook of Meta-Analysis in Ecology and Evolution*. Princeton University Press, Princeton, NJ, USA.

Larivière, V., Gingras, Y. & Archambault, É. (2009) The decline in the concentration of citations, 1900–2007. *Journal of the American Society for Information Science and Technology*, **60**, 858–862.

Magliocca, N.R., Rudel, T.K., Verburg, P.H., McConnell, W.J., Mertz, O., Gerstner, K., Heinimann, A. & Ellis, E.C. (2015) Synthesis in land change science: methodological patterns, challenges, and guidelines. *Regional Environmental Change*, **15**, 211–226.

Margulies, J.D., Magliocca, N.R., Schmill, M.D. & Ellis, E.C. (2016) Ambiguous geographies: connecting case study knowledge with global change science. *Annals of the American Association of Geographers*, **106**, 572–596.

Martin, L.J., Blossey, B. & Ellis, E. (2012) Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, **10**, 195–201.

McGill, B.J., Dornelas, M. & Field, R. (2016) A new year with a new leadership team at GEB - or how to guarantee your paper gets into GEB: Editorial. *Global Ecology and Biogeography*, **25**, 1–2.

Mengersen, K. & Gurevitch, J. (2013) Using other metrics of effect size in meta-analysis. *Handbook of Meta-Analysis in Ecology and Evolution* (eds. J. Koricheva, J. Gurevitch & K. Mengersen), pp. 72–85, Princeton University Press, Princeton, NJ, USA.

Mengersen, K., MacNeil, M.A. & Caley, M.J. (2015) The potential for meta-analysis to support decision analysis in ecology. *Research Synthesis Methods*, **6**, 111–121.

Nakagawa, S. & Cuthill, I.C. (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, **82**, 591–605.

Nakagawa, S. & Parker, T.H. (2015) Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, **13**, 88.

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

Nosek, B.A., Alter, G., Banks, G.C. *et al.* (2015) Promoting an open research culture. *Science*, **348**, 1422–1425.

Osenberg, C.W., Sarnelle, O. & Cooper, S.D. (1997) Effect size in ecological experiments: the application of biological models in meta-analysis. *The American Naturalist*, **150**, 798–812.

OSF|Tools for Transparency in Ecology and Evolution (TTEE). Available at: https://osf.io/g65cb (accessed 24 October 2016).

Parker, T.H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J.D., Chee, Y.E., Kelly, C.D., Gurevitch, J. & Nakagawa, S. (2016a) Transparency in ecology and evolution: real problems, real solutions. *Trends in Ecology & Evolution*, **31**, 711–719.

Parker, T.H., Nakagawa, S. & Gurevitch, J. (2016b) Promoting transparency in evolutionary biology and ecology. *Ecology Letters*, **19**, 726–728.

Pullin, A.S. (2012) Realising the potential of environmental data: a call for systematic review and evidence synthesis in environmental management. *Environmental Evidence*, **1**, 1.

Roche, D.G., Kruuk, L.E.B., Lanfear, R. & Binning, S.A. (2015) Public data archiving in ecology and evolution: how well are we doing? *PLOS Biology*, **13**, e1002295.

Rosenberg, M., Rothstein, H. & Gurevitch, J. (2013) Effect sizes: conventional choices and calculations. *Handbook of Meta-Analysis in Ecology and Evolution* (eds. J. Koricheva, J. Gurevitch & K. Mengersen), pp. 61–71, Princeton University Press, Princeton, NJ, USA.

Rothstein, H.R., Sutton, A.J. & Borenstein, M. (2006) *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons, Hoboken, NJ, USA.

Scheiner, S.M., Cox, S.B., Mittelbach, G.G., Osenberg, C. & Kaspari, M. (2000) Species richness, species–area curves and Simpson's paradox. *Evolutionary Ecology Research*, **2**, 791–802.

Schmidt, F.L. (1992) What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, **47**, 1173–1181.

Whittaker, R.J. (2010) Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness-productivity relationship. *Ecology*, **91**, 2522–2533.

Whittaker, R., Willis, K. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.

Zuur, A.F. & Ieno, E.N. (2016) A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, **7**, 636–645.

Received 8 November 2016; accepted 6 February 2017
Handling Editor: David Warton

S5.1.1 | **Table.** Characteristics of national forest inventories

National forest inventory	Number of plots	Survey dates	Sample plot design	Plot size (m²)
Finland	1,066	Subset of NFI 8: 1985-1986 and 1995	Cluster design, number and grid size depend on location, see text for details	100, 300
Sweden	3,863	Inventories 2005-2007 and 2008-2010	Cluster design, number and grid size depend on region. Tract size is between 300-1800 m in length.	38, 314
Germany	7,887	BWI 1 (1986-1990) and BWI 2 (2001-2002)	Cluster design, 4 subplots. Grid size depends on region. Standard grid size is 4 by 4 km	BAF 4 m^2 ha^{-1}
Wallonia	48	Inventories 1994-2003 and 2008-2011	1km by 0.5 km grid of single sample plots	63, 254, 1017
Spain	14,883	SFI 2 (1986-1996) and SFI 3 (1997-2007)	1 km by 1 km grid of single sample plots	79, 315, 707, 1964

S5.1.2 | **Table.** Characteristics of tree diversity experiments

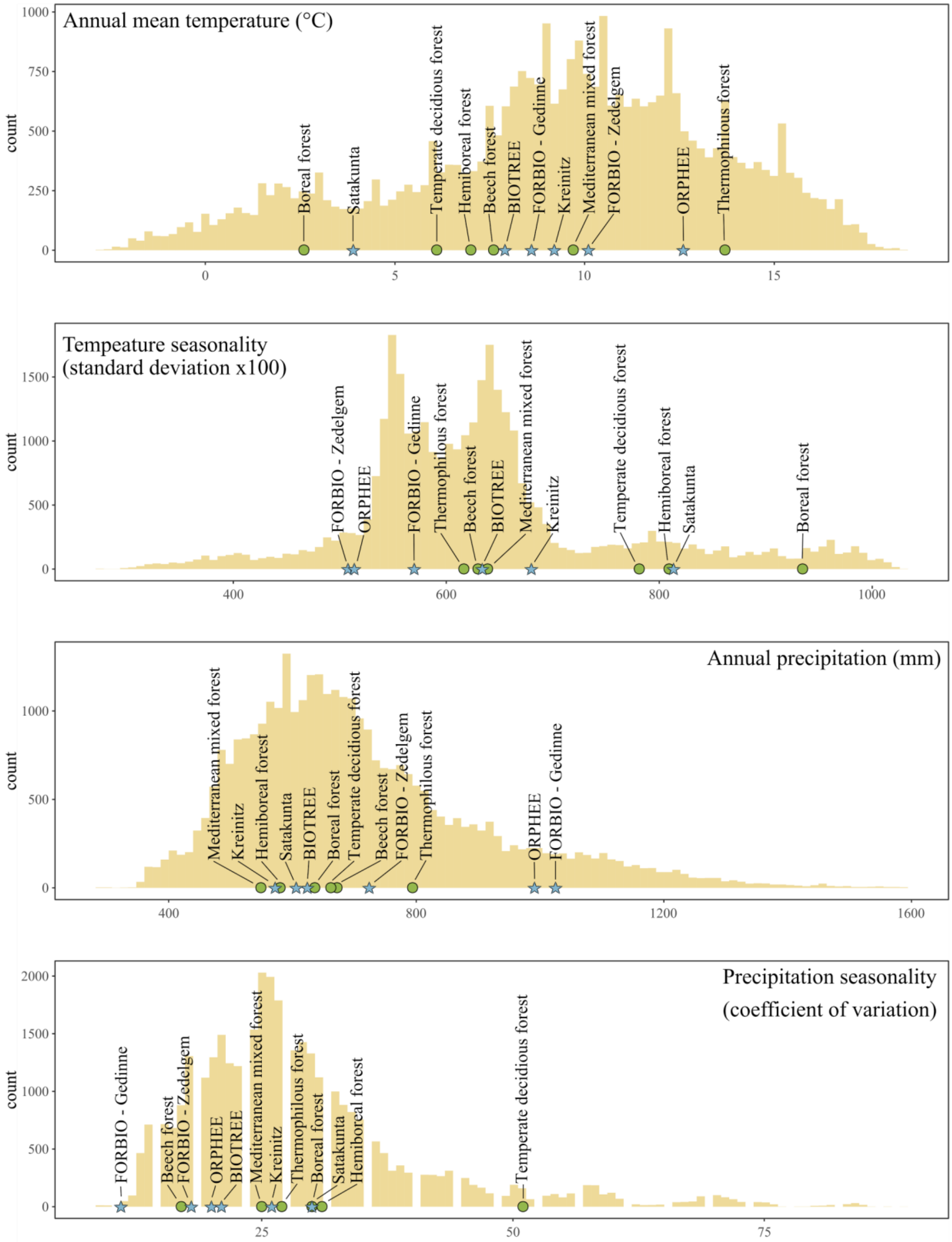
Tree diversity experiment	Country	Contact	Set-up	Growth measurement	Species richness	Tree species
Satakunta	Finland	www.sataforestdiversity.org	114 plots in 3 sites with a density of 169 trees per 400 m², planted in 1999 and partially thinned in 2013	Diameter at breast height	1, 2, 3, 5	ALGL, BESP, LASI, PIAB, PISY
Kreinitz	Germany	www.treedivnet.ugent.be/ExpKreinitz.html	96 plots in 1 site with a density of 36 trees per 25 m² plot, planted in 2000	Basal area	1, 2, 3, 5, 6	FASY, FREX, PIAB, PISY, QUSP, TICO
BIOTREE - Kaltenborn	Germany	www.biotree.uni-freiburg.de/deutsch/index.html	16 plots (each divided in 2 subplots) in 1 site with a mean density of 1944 trees per 0.6 ha plot, planted in 2004 and partially thinned	Basal area	1, 2, 3, 4	FASY, PIAB, PSME, QURO
FORBIO - Gedinne	Belgium	www.treedivbelgium.ugent.be	44 plots in 2 sites with a density of 757 trees per average plot of 1670 m², planted in 2009	Diameter at ground height	1, 2, 3, 4	ACPS, FASY, LAEU, PSME, QUPE
FORBIO - Zedelgem	Belgium	www.treedivbelgium.ugent.be	42 plots in 1 site with a density of 781 trees per 1764 m² plot, planted in 2009/2010	Diameter at ground height	1, 2, 3, 4	BESP, FASY, PISY, QURO, TICO
ORPHEE	France	www.facebook.com/orpheeexperiment	256 plots in 1 site with a density of 100 trees per 400 m² plot, planted in 2008	Tree height	1, 2, 3, 4, 5	BESP, QUPY, QURO, QUIL, PIPI

S5.1.3 | **Table** Characteristics of forest exploratories

Forest name	Country	Forest type	Number of plots	Species richness	Target tree species
North Karelia	Finland	Boreal forest	28	1, 2, 3	PIAB, PISY ,BESP
Białowieża Primeval Forest	Poland	Hemiboreal forests, nemoral coniferous and mixed broadleaved-coniferous forests	43	1, 2, 3, 4, 5	BESP , CABE, PIAB, PISY, QUSP
Hainich National Park	Germany	Beech forest	38	1, 2, 3, 4, 5	FASY, FREX, QUSP, PIAB, ACPS
Carpathian mountains	Romania	Carpathian beech forest type (temperate deciduous)	28	1, 2, 3, 4	PIAB, ABAL, FASY, ACPS
Central southern Tuscany	Italy	Thermophilous deciduous forest	36	1, 2, 3, 4, 5	QUCE, QUIL, CASA, QUSP, OSCA
Alto Tajo Natural Park	Spain	Mediterranean mixed forest	36	1, 2, 3, 4	PISY, PINI, QUFA, QUIL

S5.2 | **Table.** All tree species that occurred in any of the inventory, experimental or exploratory data sets.

ABAL - Abies alba Mill.	*ABCO - Abies concolor (Gordon) Lindl. ex Hildebr.*	*ABGR - Abies grandis*
ACCA - Acer campestre L.	*ACMO - Acer monspessulanum L.*	*ACOP - Acer opalus Mill.*
ACPL - Acer platanoides L.	*ACPS - Acer pseudoplatanus L.*	*AEHI - Aesculus hippocastanum L.*
ALGL - Alnus glutinosa (L.) Gaertn.	*ALIN - Alnus incana (L.) Moench*	*ARUN - Arbutus unedo L.*
BESP - Betula spec. - combined *B. pendula* and *B. pubescens*	*BUSE - Buxus sempervirens L.*	*CABE - Carpinus betulus L.*
CASA - Castanea sativa Mill.	*CEAT - Cedrus atlantica (Endl.) Carrière*	*CESI - Cercis siliquastrum˜ L.*
COAV - Corylus avellana L.	*CRMO - Crataegus monogyna Jacq.*	*EUCA - Eucalyptus camaldulensis Dehnh.*
FASY - Fagus sylvatica L.	*FICA - Ficus carica L.*	*FRAN - Fraxinus angustifolia Vahl*
FREX - Fraxinus excelsior L.	*ILAQ - Ilex aquifolium L.*	*JUCO - Juniperus communis L.*
JUOX - Juniperus oxycedrus L.	*JUPH - Juniperus phoenicea L.*	*JUTH - Juniperus thurifera L.*
LADE - Larix decidua Mill.	*LAEU - Larix x eurolepis*	*LAKA - Larix kaempferi (Lamb.) Carrière sec. Franco*
LASI - Larix sibirica	*MASY - Malus sylvestris Mill.*	*OLEU - Olea europaea L.*
OSCA - Ostrya carpinifolia	*PHLA - Phillyrea latifolia L.*	*PIAB - Picea abies (L.) H.Karst.*
PICA - Pinus canariensis Sweet ex Spreng.	*PICE - Pinus cembra L.*	*PIHA - Pinus halepensis Mill.*
PINI - Pinus nigra J.F.Arnold	*PIPI2 - Pinus pinea L.*	*PIPI - Pinus pinaster*
PIRA - Pinus radiata D.Don	*PISI - Picea sitchensis (Bong.) Carrière*	*PISY - Pinus sylvestris L.*
PIUN - Pinus uncinata Mill. ex Mirb.	*PITE - Pistacia terebinthus L.*	*PLOC - Platanus occidentalis L.*
POP - Populus spp.	*PONI - Populus nigra L.*	*POTR - Populus tremula L.*
PRU - Prunus spp.	*PRAV - Prunus avium L.*	*PRDO - Prunus domestica˜L.*
PRMA - Prunus mahaleb L.	*PRPA - Prunus padus L.*	*PSME - Pseudotsuga menziesii (Mirb.) Franco*
PYCO - Pyrus communis L.	*QUE - Quercus spp.*	*QUCA - Quercus canariensis Willd.*
QUCE - Quercus cerris	*QUFA - Quercus faginea Lam.*	*QUIL - Quercus ilex L.*
QUPY - Quercus pyrenaica Willd.	*QURO - Quercus robur L.*	*QURU - Quercus rubra L.*
QUSP - Quercus spec. - combines Q. petraea and Q. pubescens Willd. (Q. Humilis)	*QUSU - Quercus suber L.*	*ROPS - Robinia pseudacacia L.*
SAAL - Salix alba L.	*SAAT - Salix atrocinerea Brot.*	*SACA - Salix caprea L.*
SACI - Salix cinerea L.	*SANI - Sambucus nigra L.*	*SOAR - Sorbus aria (L.) Crantz*
SOAU - Sorbus aucuparia L.	*SODO - Sorbus domestica L.*	*SOTO - Sorbus torminalis (L.) Crantz*
TABA - Taxus baccata L.	*TICO - Tilia cordata*	*TIL - Tilia spp.*
TIPL - Tilia platyphyllos Scop.	*ULGL - Ulmus glabra Huds.*	*ULMI - Ulmus minor Mill.*



S5.3 | **Figure.** Climate conditions across the three research approaches (extracted from the WorldClim dataset, Hijmans *et al.* 2008). Histograms of the forest inventory plots are shown in the background in yellow. Tree diversity experiments are marked with star symbols and forest exploratories are marked with circles.

S5.4 | Table. Species compositions that are shared between the three different research approaches (Inv - inventory approach, Exper - experimental approach, Explor - exploratory approach). Colouring indicates if the respective tree species composition is represented in the two compared approaches.

Species composition	NFIs / Exper	NFIs / Explor	Exper / Explor
ABAL	no	yes	no
ABAL ACPS FASY	no	yes	no
ABAL ACPS FASY PIAB	no	yes	no
ABAL ACPS PIAB	no	yes	no
ABAL FASY	no	yes	no
ABAL FASY PIAB	no	yes	no
ABAL PIAB	no	yes	no
ACPS	yes	yes	yes
ACPS FASY	yes	yes	yes
ACPS FASY FREX	no	yes	no
ACPS FASY FREX PIAB	no	yes	no
ACPS FASY PIAB	no	yes	no
ACPS FASY QUSP	no	yes	no
ACPS FREX	no	yes	no
ACPS PIAB	no	yes	no
ACPS PSME	yes	no	no
ALGL	yes	no	no
ALGL BESP	yes	no	no
ALGL BESP PISY	yes	no	no
ALGL PIAB	yes	no	no
BESP	yes	yes	yes
BESP FASY PISY	yes	no	no
BESP QURO	yes	yes	yes
CABE	no	yes	no
CABE PIAB	no	yes	no
CABE PIAB QURO	no	yes	no
CABE PISY	no	yes	no
CABE PISY QURO	no	yes	no
CABE QURO	no	yes	no
CASA	no	yes	no
CASA QUIL	no	yes	no
CASA QUIL QUSP	no	yes	no
CASA QUSP	no	yes	no
FASY	yes	yes	yes
FASY FREX	yes	yes	yes
FASY FREX PIAB	yes	yes	yes
FASY FREX QUSP	no	yes	no
FASY PIAB	yes	yes	yes
FASY PIAB PISY	yes	no	no
FASY PIAB PSME	yes	no	no
FASY PIAB PSME QURO	yes	no	no
FASY PIAB QURO	yes	no	no
FASY PIAB QUSP	no	yes	no
FASY PISY	yes	no	no
PIAB PSME	yes	no	no
PIAB PSME QURO	yes	no	no
PIAB QURO	yes	yes	yes
PINI	no	yes	no
PINI PISY	no	yes	no
PINI PISY QUFA	no	yes	no
PINI PISY QUFA QUIL	no	yes	no
PINI QUFA	no	yes	no
PINI QUFA QUIL	no	yes	no
PINI QUIL	no	yes	no
PIPI2	yes	no	no
PIPI2 QUIL	yes	no	no
PIPI2 QUIL QUPY	yes	no	no
PIPI2 QUPY	yes	no	no
PIPI2 QUPY QURO	yes	no	no
PIPI2 QURO	yes	no	no
PISY	yes	yes	yes
PISY QUFA	no	yes	no
PISY QURO	no	yes	no
PSME	yes	no	no
PSME QURO	yes	no	no
QUFA	no	yes	no

BESP FASY PISY QURO	yes	no	no
BESP PIAB	yes	yes	yes
BESP PIAB PISY	yes	yes	yes
BESP PIAB PISY QURO	no	yes	no
BESP PIAB QURO	no	yes	no
BESP PIPI2	yes	no	no
BESP PIPI2 QUPY QURO	yes	no	no
BESP PIPI2 QURO	yes	no	no
BESP PISY	yes	yes	yes
BESP PISY QURO	no	yes	no
BESP QUPY	yes	no	no
BESP QUPY QURO	yes	no	no
FASY PISY QURO	yes	no	no
FASY PSME	yes	no	no
FASY QURO	yes	no	no
FASY QUSP	no	yes	no
FASY TICO	yes	no	no
FREX	yes	yes	yes
FREX PIAB	yes	yes	yes
FREX PIAB PISY	yes	no	no
FREX PISY	yes	no	no
PIAB	yes	yes	yes
PIAB PISY	yes	yes	yes
PIAB PISY QURO	No	yes	no
QUFA QUIL	no	yes	no
QUIL	yes	yes	yes
QUIL QUPY	yes	no	no
QUIL QURO	yes	no	no
QUIL QUSP	no	yes	no
QUPY	yes	no	no
QUPY QURO	yes	no	no
QURO	yes	yes	yes
QURO QUSP	no	yes	no
QUSP	no	yes	no

S5.5 | **Text.** Sampling design of national forest inventory plots.

Data from the national forest inventories (NFIs) of Finland, Sweden, Germany, Wallonia and Spain were compiled as part of the Inventory Platform of FunDivEUROPE. To select comparable data from the different inventories, only those trees with a diameter at breast height (dbh) of 10 cm and plots with consecutive surveys and no indication of harvest between survey dates were included in the analyses (see Baeten *et al.* 2013 and Ratcliffe *et al.* 2016 for more details).

FINLAND

The Finnish data is a subset of permanent sample plots established to follow changes in forest vegetation. Data included in this study is from two surveys: 1985 to 1986 and 1995. The sample plots are located on forest land in a systematic grid across the country (Mäkipää & Heikkinen 2003; Tomppo & Tuomainen 2009) forming a regular network of clusters. The size of the grid and the number of plots within each cluster depends on the location. In Southern Finland, the grid is 16 by 16 km square with four plots in each cluster at 400 m intervals. In Northern Finland, the grid is a 24 by 32 km rectangle with three plots per cluster at 600 m intervals.

The plot size depends on the dbh of the sample trees: 100 m^2 if trees < 10.5 cm in dbh, and 300 m^2 if trees >10.5 cm in dbh. All plots are located in intensively managed forests, in which suppressed trees were thinned.

SWEDEN

We received data from the permanent sample tracts of the Swedish NFI. The inventory uses a randomly planned regular sampling grid and includes about 4,500 permanent tracts, each surveyed every five years (Fridman *et al.* 2014). Plots in the first census were surveyed between 2003 and 2005 and plots in the second census were surveyed between 2008 and 2010. The tracts are rectangular and have different dimensions depending on the location within the country. Each tract has between 4 and 8 circular sample plots.

Trees greater than 1.3 m high are sampled in two different plot sizes depending on the dbh of the tree: 40-99 mm dbh: 3.5 m radius; and greater than 100 mm dbh: 10 m radius.

GERMANY

We received data from the first two German NFIs. The first inventory was surveyed between 1986 and 1990 (undertaken in West Germany only) and the second inventory was surveyed between 2001 and 2002. The mean survey period was 12 years.

The German NFI is based on a systematic rectangular grid, the dimensions of which are determined by the Federal State; the standard size is 4 by 4 km and it is intensified in some States to either 2.83 by 2.83 km or 2 by 2 km (Polley *et al.* 2009). In each grid square is a quadratic tract of 150 m in length. Each corner of the tract has a sample plot and the tracts are surveyed if at least one of the corners is in forest.

Trees with a minimum dbh of 10 cm, in the first inventory, and 7 cm, in the second inventory, were surveyed based on callipered angle count sampling using a basal area factor of 4 m^2 ha^{-1}.

WALLONIA

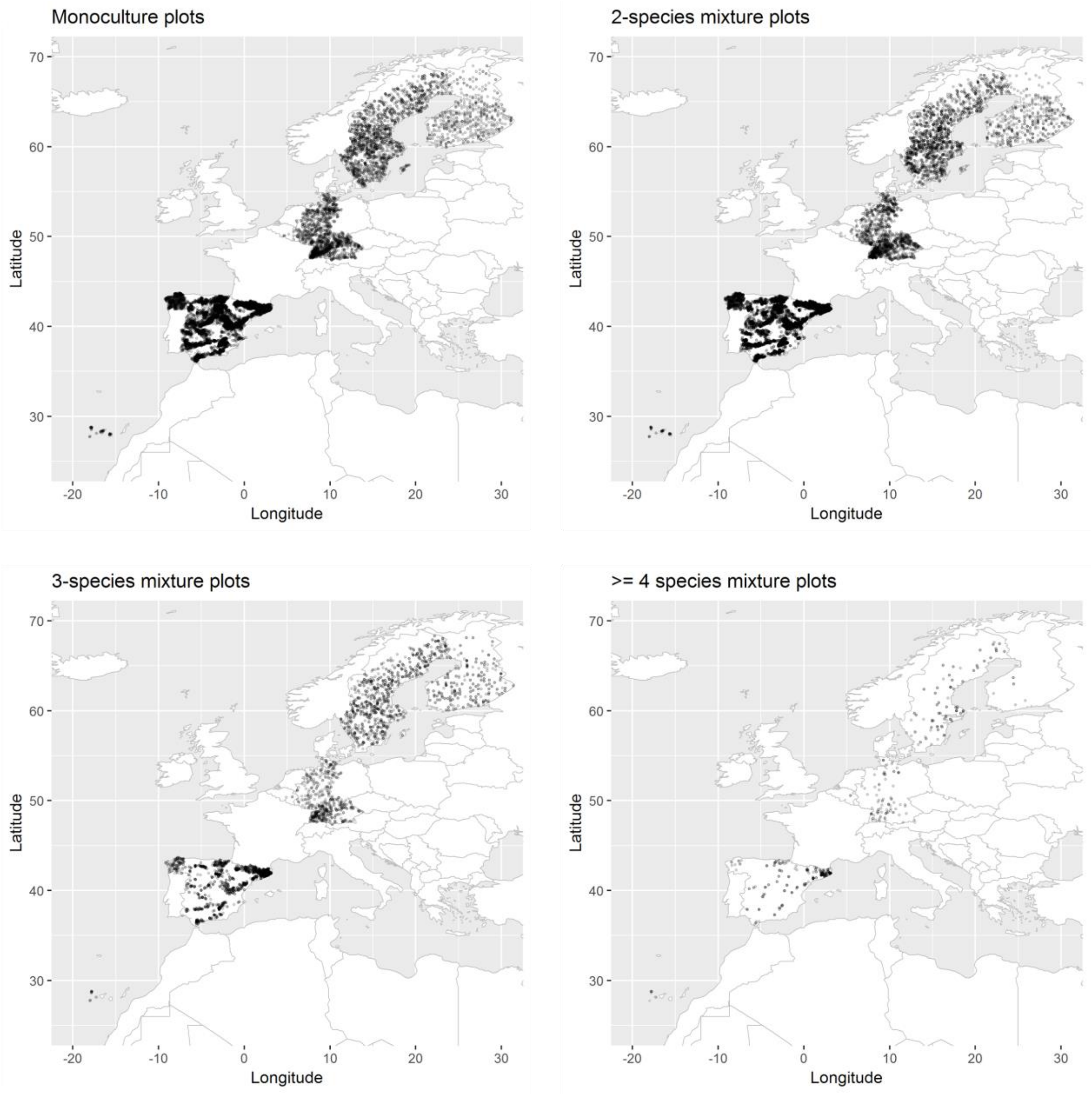
The Walloon NFI follows a systematic non-stratified sampling methodology on a 1 km by 0.5 km grid (Rondeux & Wagner 2009). One circular sampling plot is located within each grid intersection. Areas are sampled if the area of land is greater than 0.1 ha and has at least 10% covered by a forest canopy (trees must be able to reach a minimum of 5 m). Plots in the first census were surveyed between 1994 and 2003 and plots in the second census were surveyed between 2008 and 2011.

The inventory employs a variable plot size depending on the circumference of the tree: Trees with a circumference between 20-69 cm: 4.5 m radius; 70-119 cm: 9 m radius; and greater than 120 cm: 18 m radius.

SPAIN

We used data from the permanent sample plots of the second and third Spanish NFIs. Plots were surveyed between 1986 and 1996 and between 1997 and 2007, with a mean survey period of 10 years. The sample plots of the Spanish NFI are on a systematic 1 km^2 grid in forested areas of the country, and are not grouped in tracts but simply one plot in each grid square (Villaescusa & Díaz 1998; Villanueva 2005).

The Spanish NFI used a variable radius plot size depending on the dbh of the sample trees; each plot has four nested subplots of 5, 10, 15 and 25 m radius and the minimum dbh for a tree to be recorded within a subplot is 7.5 cm, 12.5 cm, 22.5 cm and 42.4 cm, respectively.



S5.6. | **Figure.** Locations of the mixed and monospecific plots (altogether 16,773 plots) retained in the inventory approach after the omission the mono-mix comparisons above the 10 percent percentile (see Figure S4.9).

S5.7 | **Appendix.** Tree growth measurements in the exploratory platform.

In the exploratory platform tree growth was estimated as yearly radial stem increment (mm year^{-1}) between 1990 and 2010 as measured from wood cores that were extracted at breast height between March and October of 2012. In each plot these wood cores were taken from 12 trees in monospecific and six trees per species in mixture plots (except in Poland where only five cores per species were taken in each plot). Each sample was later cross dated against an averaged reference curve obtained from all chronologies that were measured on the same species at the same site. After excluding samples with poor agreement we ended up with 2926 tree ring chronologies and omitted two plots in Italy due to lack of data. The process of radial growth measurements is described in detail in (Jucker *et al.* 2014).

S5.8 | **Appendix.** Data preparation to compare the effect of species mixing across all research approaches.

1. SUBSET PLOTS

Inventories	Retained only those plots with multiple survey dates and that were neither harvested nor managed in between subsequent surveys.
Experiments	-
Exploratories	-

2. CALCULATE SPECIES GROWTH ESTIMATES AND PLOT CHARACTERISTICS

Inventories	*Species growth estimate*: average increment in basal area ha^{-1} year^{-1} basal area (2nd survey)$^{-ha}$ for each focal tree species *Species proportions*: summed basal area ha^{-1} of the focal species divided by summed basal area ha^{-1} of all trees (2nd survey) *Community density*: sum of basal area ha^{-1} of all trees (2nd survey) *Community heterogeneity*: mean basal area of all trees divided by the respective standard deviation (2nd survey) *Climate variables*: Mean annual temperature, temperature seasonality and annual precipitation were extracted from the WorldClim dataset *Forest type*: Based on the EEA Technical Report 9
Experiments	*Species growth estimates*: mean diameter at ground or breast height, tree height or basal area basal area ha^{-1} per species in each plot *Species proportions*: not accounted since species were planted in equal proportions *Community density*: - not available *Community heterogeneity*: - not available
Exploratories	*Species growth estimates*: average increment in basal area (S7) ha^{-1} year^{-1} basal area^{-1} for each focal tree species *Species proportions*: summed basal area ha^{-1} of the focal species divided by summed basal area ha^{-1} of all trees *Community density*: sum of basal area ha^{-1} of all trees *Community heterogeneity*: mean basal area of all trees divided by the respective standard deviation *Climate variables*: Mean annual temperature, temperature seasonality and annual precipitation were extracted from the WorldClim dataset *Forest type*: Based on the EEA Technical Report 9

3. DETERMINE SPECIES RICHNESS PER PLOT

Inventories	Species richness levels were assigned based on species proportions:	
	1: One dominant species > 90%;	27,975 plots
	2: Two dominant species, together > 90%;	14,478 plots
	3: Three dominant species, together > 90%;	4,436 plots
	Higher species richness3: > Three dominant species, together > 90%;	865 plots
	Any non-focal/non-dominant species must not exceed 10%.	
	Plots that did not meet these criteria were filtered out.	
Experiments	-	
Exploratories	-	

4. MATCH SIMILAR MIXED AND MONOSPECIFIC PLOTS

Inventories

We applied the following procedure to assign, for each tree species within each forest type the species growth in monospecific and mixed plots that should be as similar as possible with regard to the community density, heterogeneity and climate conditions.
For each plot the values of community density, heterogeneity and climate conditions were standardized (divided by mean and standard deviation).
Based on these standardized values we calculated the Euclidean distance between all plots in which this species occurred within the respective forest type.

- We applied a nearest neighbour algorithm that assigned pairs of monospecific and mixed plots that were most similar (i.e. with the lowest Euclidean distance) while at the same time minimizing the summed Euclidean distance between all pairs (based on a binomial generalized linear model, R-package *MatchIt*, *(*Ho *et al.* 2007, 2011*)*. The histogram of all Euclidean distances is shown in Figure S4.9.
- Assigned pairs with a distance above the 90% percentile were omitted to not compare very different mixed and monospecific plots (see Figure S4.9).

Experiments

Tree species in mixed and monospecific plots are, by design, growing in very comparable conditions.

Exploratories

Tree species in mixed and monospecific plots are, by design, growing in very comparable conditions.

5. CALCULATE GROWTH SUMMARY STATISTICS

Inventories

- Within each forest type and each national forest inventory we calculated for each species:
- Mean and standard deviation of species growth together with the number of plots in monospecific, two-, 3- and higher species richness mixtures.

Experiments

- Within each tree diversity experiment we calculated for each species:
- Mean and standard deviation of species growth together with the number of plots in monospecific, two-, 3- and higher species richness mixtures.

Exploratories

- Within each forest type we calculated for each species:
- Mean and standard deviation of species growth together with the number of plots in monospecific, two-, 3- and higher species richness mixtures.

6. CALCULATE EFFECT SIZES

Inventories

- Within each forest type we calculated for each species:
- The log response ratio from the contrasting species growth in all mixed vs. all monospecific plots
- These effect sizes were calculated separately for each level of species richness (two-, three- and higher species mixtures). The same set of monospecific plots was thereby contrasted with different sets of mixture plots (representing the different levels of species richness)

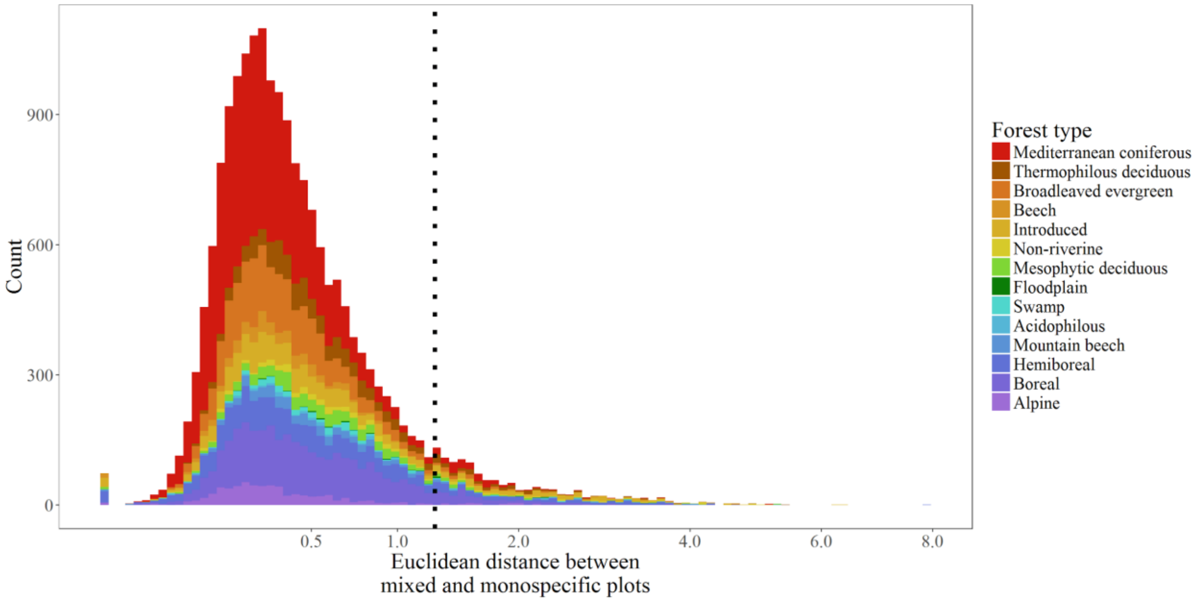
Experiments

- Within each experiment we calculated for each species:
- The log response ratio from the contrasting species growth in all mixed vs. all monospecific plots
- These effect sizes were calculated separately for each level of species richness (two-, three- and higher species mixtures). The same set of monospecific plots was thereby contrasted with different sets of mixture plots (representing the different levels of species richness)

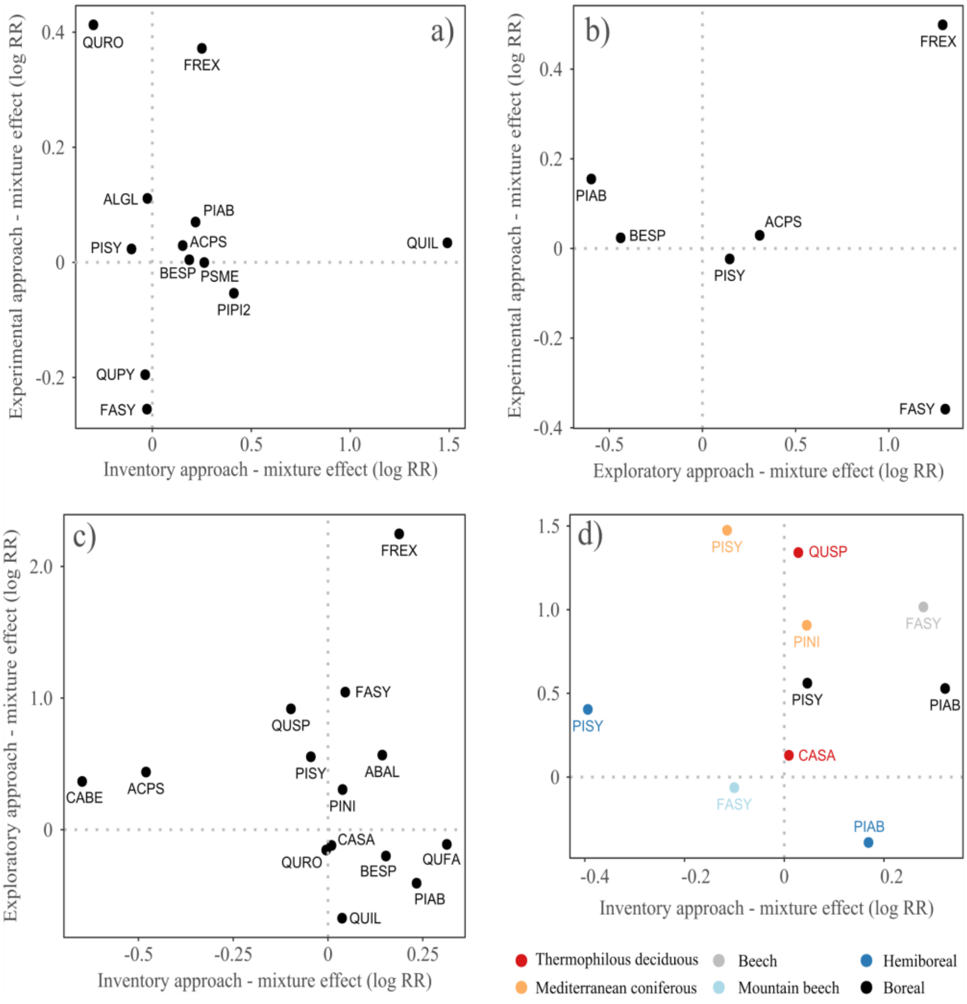
Exploratories

- Within each forest type we calculated for each species:
- The log response ratio from the contrasting species growth in all mixed vs. all monospecific plots
- These effect sizes were calculated separately for each level of species richness (two-, three- and higher species mixtures). The same set of monospecific plots was thereby contrasted with different sets of mixture plots (representing the different levels of species richness)

Log response ratios were calculated as $lnR\ = ln\ (\frac{\hat{x}_{mix}}{\hat{x}_{mono}})$



S5.9 | **Figure.** Histogram of Euclidean distances between assigned mixed and monospecific plots in the inventory approach; calculated from standardized values of mean annual temperature, temperature seasonality, annual precipitation, precipitation seasonality, slope of the plot and the sum and coefficient of variation of the basal area of all tree individuals (m² ha^{-1}). The dotted line indicates the 90% percentile above which mixed-monoculture-comparisons were omitted due to the high differences in plot conditions.



S5.10 | Figure. Comparison of tree species mean effect sizes (log response ratios) of growth in mixed compared to monospecific plots obtained from three different research approaches (experimental, exploratory and inventory approach). Depicted are the mean effect sizes of only those species, species compositions and forest types that were shared between the compared research approaches (a: inventories versus experiment, b: exploratories versus experiments, c: inventories versus exploratories, d: inventories versus exploratories when species responses were separated by forest type). Abbreviations: ABAL: *Abies alba* Mill., ACPS: *Acer pseudoplatanus* L., BESP: *Betula spec*., ALGL: *Alnus glutinosa* (L.) Gaertn., CABE: *Carpinus betulus* L., CASA: *Castanea sativa* Mill., FASY: *Fagus sylvatica* L., FREX: *Fraxinus excelsior* L., PIAB: *Picea abies* (L.) H.Karst., PINI: *Pinus nigra* J.F.Arnold, PIPI2: *Pinus pinea* L., PISY: *Pinus sylvestris* L., PSME: *Pseudotsuga menziesii* (Mirb.) Franco, QUFA: *Quercus faginea* Lam., QUIL: *Quercus ilex* L., QUPY: *Quercus pyrenaica* Willd., QURO: *Quercus robur* L., QUSP: *Quercus spec* - combines *Q. petraea* and *Q. pubescens Willd. (Q. Humilis)* (Table S2).

S.6.1 | **Table.** Prisma checklist.

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	-
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5-6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5-6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7

Section/topic	#	Checklist item	Reported on page #
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	-
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8,9
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8,9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8,9
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	supplement
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	supplement
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	-
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 2, supplement
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	9-14
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	supplement

Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	9-14, supplement
Section/topic	#	Checklist item	Reported on page #
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15-16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15-17
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	18

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

S6.2 | **Text.** Research Strategy and Inclusion Criteria

S5.2.1. Search Strategy

On August 1st, 2016 we searched the following databases for relevant literature:

Thomas Reuters Web of Knowledge

We retrieved all literature listed for the following search term:

- "forest" AND "tree" AND (herbivor* OR pest*) AND (diversit* OR richness OR monocultur* OR mixtur* OR polycultur* OR plantatio*)

Google Scholar

The search was performed with the program *Publish or Perish 4*.

We retrieved the first 1000 entries from the following combinations of search terms:

- All of the words: *Forest Tree Diversity Herbivory*
- Any of the words: *Richness Damage Rate Monocultures Mixtures Polycultures*

- All of the words: *Forest Tree Monocultures Mixtures Herbivores*
- Any of the words: *Diversity Damage Rate Polycultures*

- All of the words: *Forest Herbivore Resistance Diversity Tree Richness*
- Any of the words: *Monoculture Plantations Mixture Insects*

Cabdirect

We retrieved all entries from the following combinations of search terms:

- *forest AND tree AND herbivor* AND diversity OR richness OR monocultur* OR mixtur* OR polycultur* OR plantatio* OR damage*

S.5.2.2. Study Inclusion Criteria

We included all studies conducted in forests or forest plantations that reported either the Shannon/Simpson diversity, species richness or a comparison of monocultures and mixtures of tree species per plot together with the corresponding information on one of the following aspects of insect herbivory: damage inflicted on trees, abundance/density on trees or in traps, incidence rate (proportion of trees or tree parts attacked or occupied by herbivores) or the species richness of herbivores collected on trees or in traps.

Types of Intervention

We considered studies that i) manipulated tree species diversity in experimental plantings, ii) compared monocultures with stands of higher tree species richness or iii) compared proximal forest stands of varying tree species diversity. Studies that compared tree stands with varying densities of trees planted were not included into the analyses for not confounding any diversity with density effects.

Types of Comparator

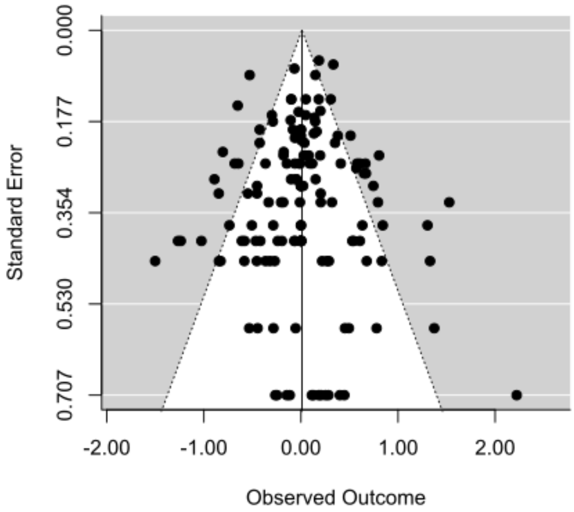
i) Damage on trees inflicted by insect herbivores
ii) Abundance/density of herbivorous insect species
iii) Incidence rate of herbivorous insect species
iii) Species richness of herbivorous insect species

S.5.2.3. Assessment of Study Relevance

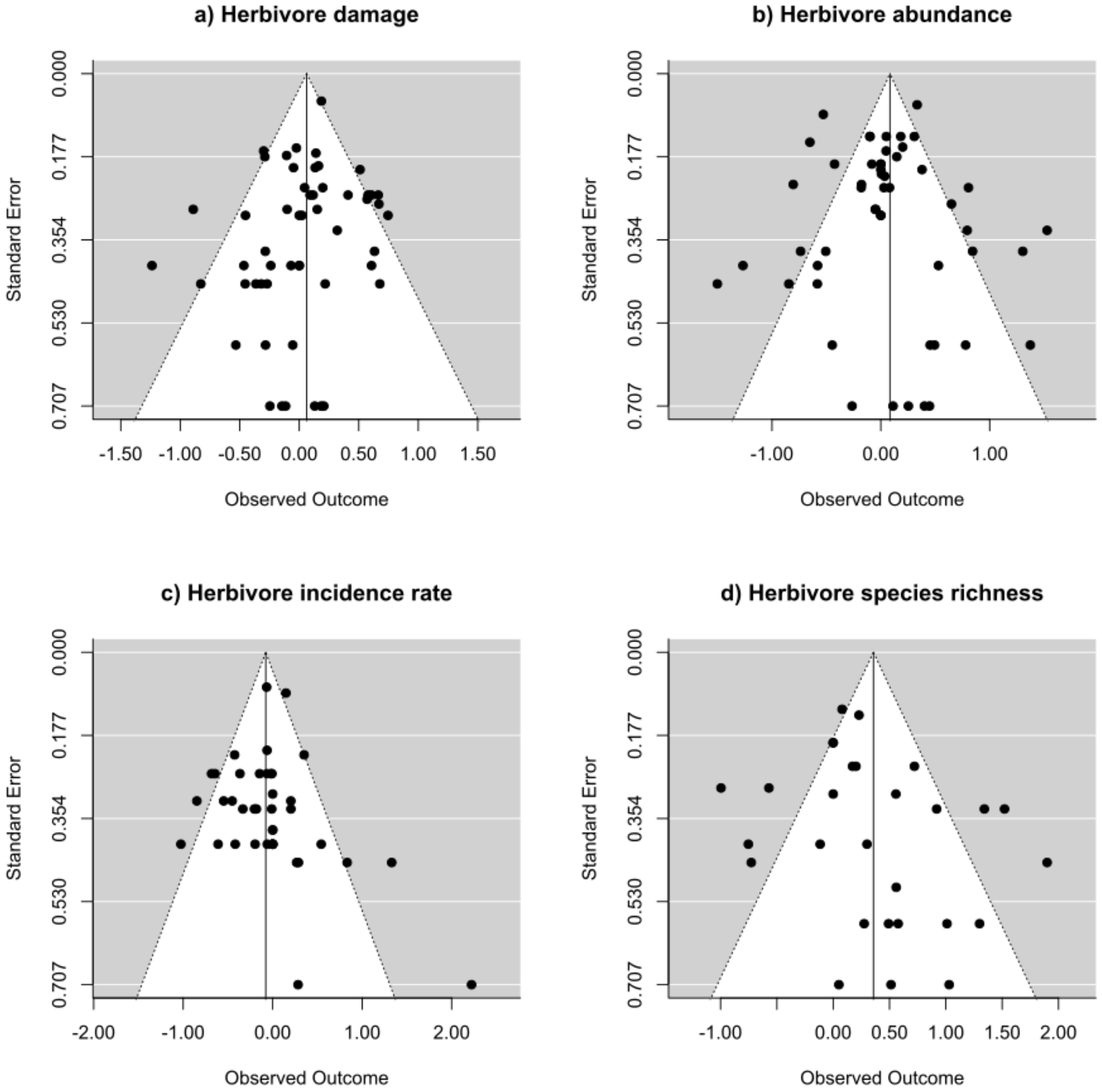
Studies were not included into the analysis for the following reasons:

- ➔ No established forest
    - o Studies were not conducted in tree stands but rather on single individuals, botanical gardens, mixed landscapes, etc.
- ➔ No comparison of tree diversity levels
    - o Compared sites had the same diversity of tree species
- ➔ No herbivory reported
    - o Study did report neither of the four focal aspects of insect herbivory
- ➔ Not a primary study
    - o Reviews, opinion papers, conceptual papers, simulation models, etc… that did not report experimental or observational data
- ➔ Paper not found/no access
- ➔ Manipulation of herbivore pressure
    - o Studies manipulated herbivory and assessed consequences on tree diversity
- ➔ Focus on soil interactions

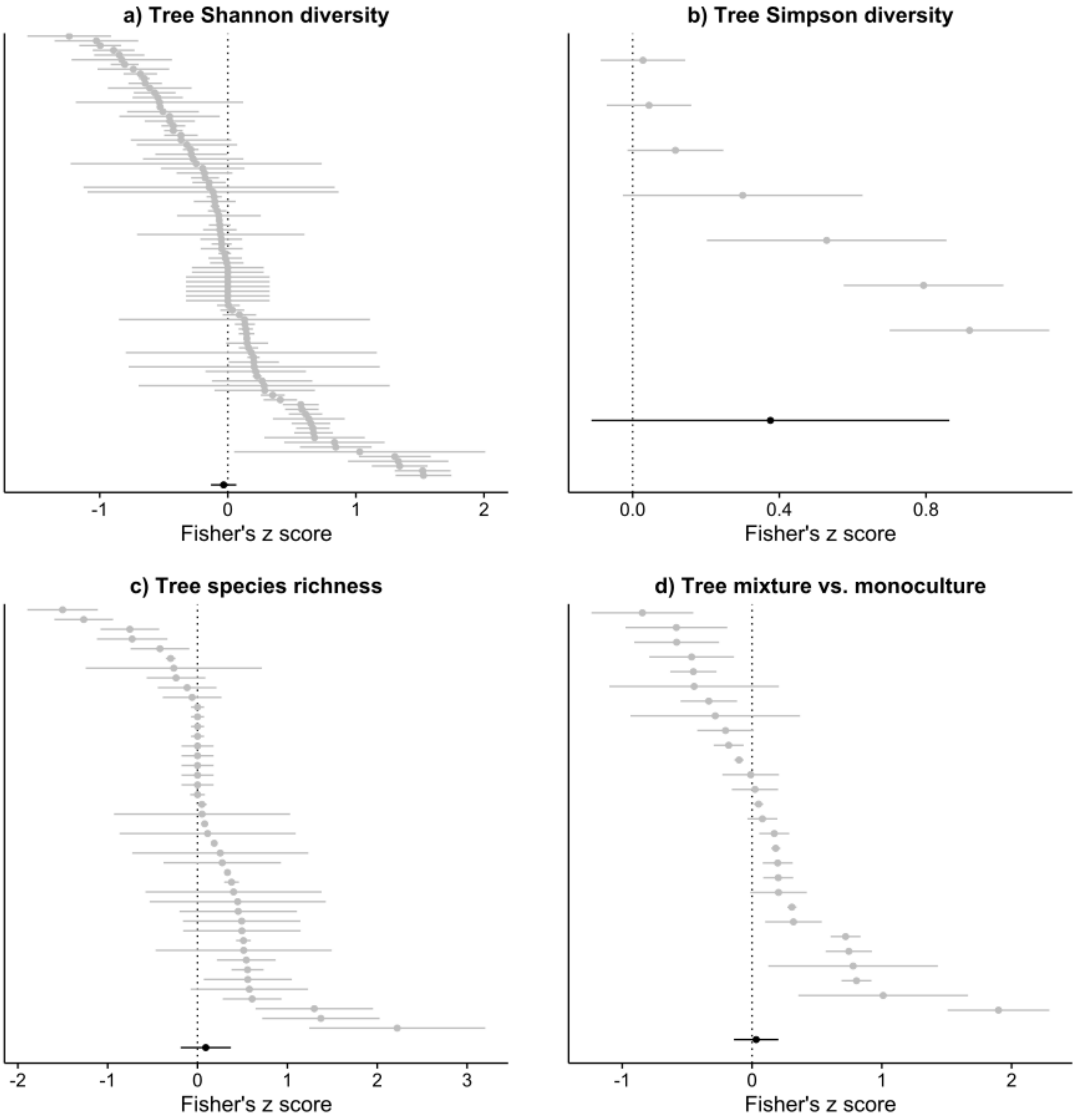
- Unclear herbivory measurement
    - It was not possible to assign herbivory to either damage, abundance, incidence rate or species richness
- More suitable study on the same site
    - A more recent or a more comprehensive study was published from the same sites
- Forest plots are not comparable
    - Forest plots differ in land-use, succession status, environmental conditions, tree density or fragment size so that comparisons of associational effects are likely flawed
- Seed survival reported



S6.4.1 | **Figure.** Funnel plot for the combined response different measures of insect herbivory (damage, abundance and incidence rate) to an increase in tree diversity. Observed outcome refers to transformed correlation coefficients (Fisher's *z*-scores). The vertical line represents the grand mean effect size which is not significantly different from zero.



S6.4.2 | **Figure.** Funnel plots for the response of different measures of insect herbivory to an increase in tree diversity. Observed outcome refers to transformed correlation coefficients (Fisher's *z*-scores). The vertical line represents the grand mean effect size which is statistically significant from zero for the response of herbivore abundance and species richness.

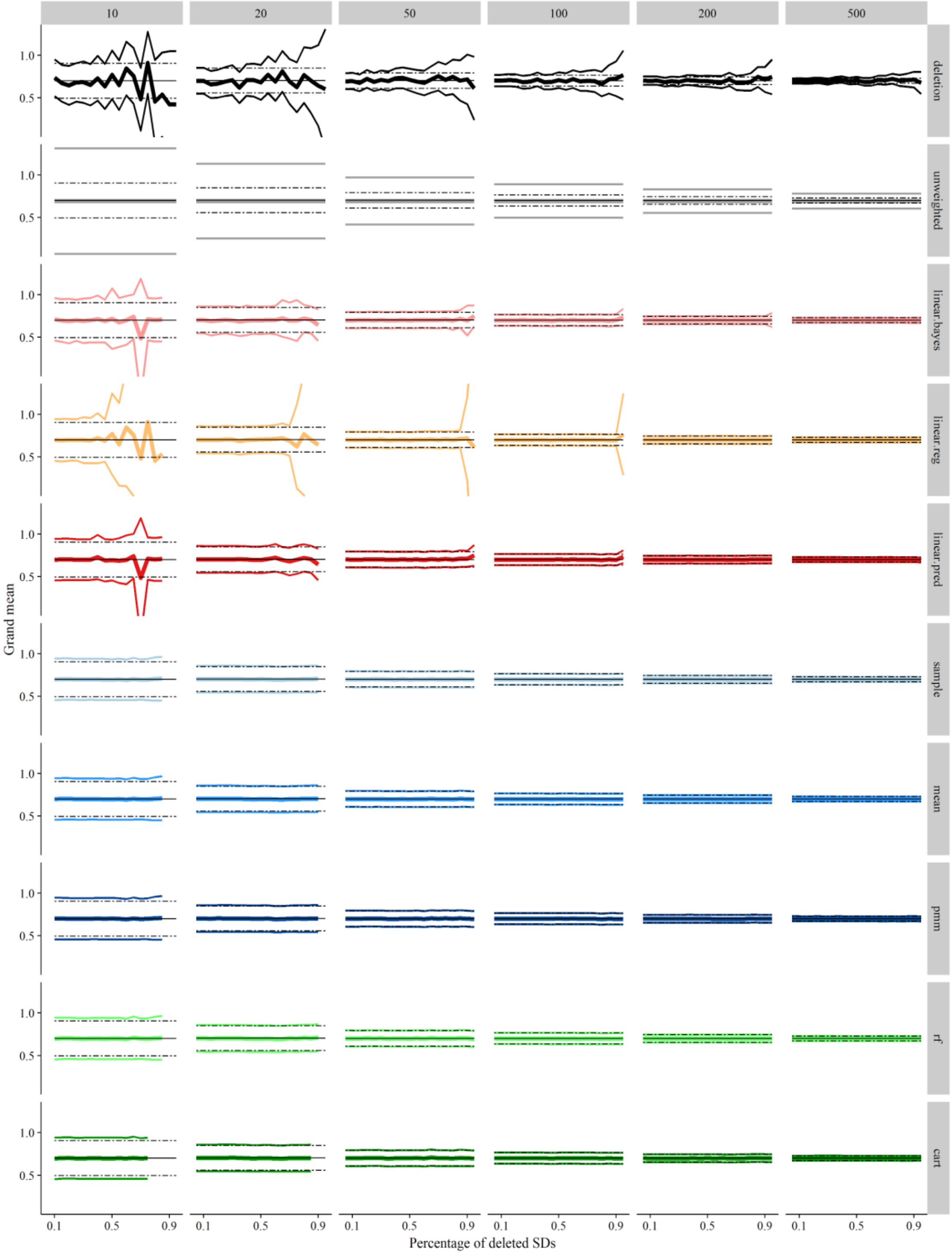


S6.4.3 | Figure. Forest plot for the transformed correlation coefficients (Fisher's *z*-scores) between insect herbivory (damage, abundance and incidence rate combined) and a) the Shannon diversity, b) the Simpson diversity, c) the species richness and d) a single and mixed stands of forest trees. Each point represents the mean Fisher's *z*-score and the approximated confidence interval (= mean ± standard error * 1.96) for a single study case. Negative values indicate associational resistance while positive values indicate associational susceptibility. Grand mean effect sizes together with their 95% bootstrap confidence intervals are shown at the bottom of each forest plot.

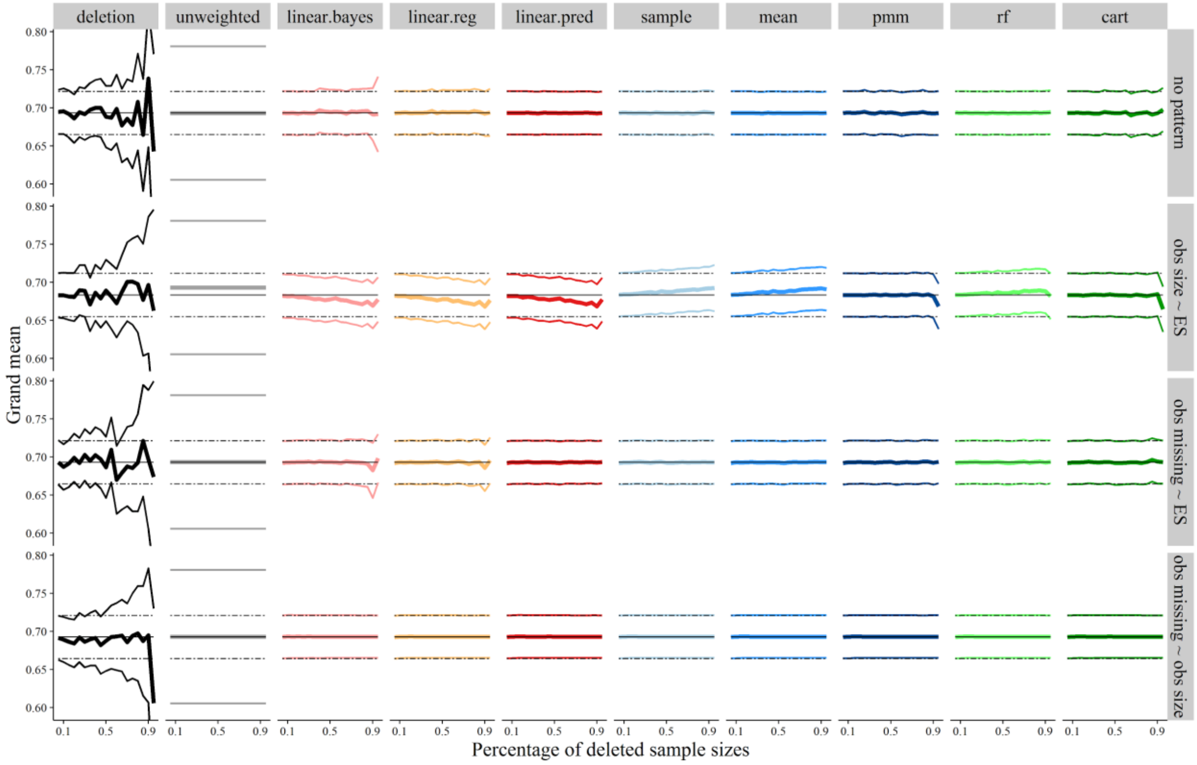
S7.2 | **Text.** Review of published meta-analyses

On September 13, 2015 we executed a search query in the Web of Science with the search term “meta-analysis AND ecology”. The obtained 483 studies were screened to fulfil the following criteria: i) the publication was based on data (i.e. excluding conceptual and methodological articles), ii) the research field was ecology, iii) the research was original (i.e. excluding re-analyses of previously published meta-analyses), iv) the meta-analysis was based on primary literature (i.e. excluding studies based on raw data). Seven articles were omitted because we could not gain access.

This screening resulted in 178 publications which were read in detail to answer the following questions: 1) Did the authors note that primary studies were missing variances? 2) How did the authors deal with these missing variances?



S7.3 | Figure. The effects of multiple imputation of missing standard deviations on the estimated grand mean under the consideration of dataset size (10, 20, 50, 100, 200 and 500 study cases) and the proportion of deleted SDs (5-95%, in steps of 5%). Imputation methods are listed in Table 2 and shown on the right. Solid and dotted lines in black show the estimated grand mean effect and the corresponding approximated 95% confidence interval from the full dataset (i.e. with all standard deviations). Coloured lines show the grand means and confidence interval obtained from the imputed data sets.



S7.4 | Figure. Grand means obtained after multiple imputation of missing sample sizes (obs) when effect sizes were weighted by the approximated variance measure (eqn 7.3) under the simulation that sample sizes were randomly distributed and deleted ("no pattern"), sample sizes correlated with the corresponding effect sizes (obs size ~ ES), the probability of deleting sample sizes negatively correlated with the corresponding effect sizes (obs missing ~ ES) or with the inverse of the sample sizes (obs missing ~ obs size). In every artificial dataset (with 500 study cases) between 5% and 95% of the sample sizes in the treatment group were deleted (in steps of 5%) and then imputed with the methods described in Table 7.2. Black solid and dotted lines show the grand mean and confidence interval of the full dataset (i.e. with all sample sizes available). Coloured lines show the estimated grand means and confidence intervals obtained from the imputed data sets.

S7.5 | **Text.** Imputation and re-analysis of a published dataset

Marczak, Thompson & Richardson (2007) compiled 115 datasets from 32 studies to investigate the response of consumer biomass and density to resource subsidies across recipient habitat types, trophic levels and functional groups. The dataset is publicly available and, in contrast to most of the reviewed meta-analyses, the authors performed an unweighted analysis and still provided all available standard deviations (SDs, 94 out of 116 study cases).

Prior to the imputation of missing SDs we determined that the mean of effect sizes with reported SDs did not differ from the mean of incompletely reported effect sizes (22% of effect sizes without SDs, Student's t-test = 0.7, df = 18.3, p = 0.5). Assuming that chances of SDs to not be reported did not correlate with effect sizes we decided to impute missing SDs via predictive mean matching (method *pmm*) rather than any method that is based on linear regression (*linear.bayes*, *linear.reg* and *linear.pred*) because the latter might impute SD values that are negative.

We re-analysed the dataset with a linear model that accounted for the variation in effect sizes between study cases (i.e. a so-called random-effects meta-analysis) with the *rma* function of the *metafor* package in R (Viechtbauer 2010; R Core Team 2013). Thereby, we did not obtain the same grand mean effect size as in the original publication of Marczak, Thompson & Richardson (2007) but aimed at demonstrating the effect of performing i) an unweighted analysis (method unweighted), ii) a weighted analysis that omits incompletely reported studies (method deletion) and iii) a weighted analysis with missing SDs filled by the mean value of the reported SDs (method *mean*) and iv) a weighted analysis with missing SDs imputed via predictive mean matching (method *pmm*). We calculated the grand mean and bootstrap-corrected 95% confidence interval for each imputation method (Fig. 7.7).

Referring to Figure 1 in the original publication of Marczak, Thompson & Richardson (2007), we then tested the effect of habitat type by including this variable as a fixed effect in the aforementioned random-effects model that were based on the dataset with (i) omitted incompletely reported studies and (i) SDs that were imputed via predictive mean matching (method *pmm*). The obtained group mean effect sizes and corresponding confidence intervals are shown in Figure 7.7.

9.2 | References

Adams, J.M. & Zhang, Y. (2009) Is there more insect folivory in warmer temperate climates? A latitudinal comparison of insect folivory in eastern North America. *Journal of Ecology*, **97**, 933–940.

Adams, J.M., Zhang, Y., Basri, M., & Shukor, N. (2009) Do tropical forest leaves suffer more insect herbivory? A comparison of tropical versus temperate herbivory, estimated from leaf litter. *Ecological Research*, **24**, 1381–1392.

Adams, W.M., Aveling, R., Brockington, D., Dickson, B., Elliott, J., Hutton, J., Roe, D., Vira, B., & Wolmer, W. (2004) Biodiversity conservation and the eradication of poverty. *Science*, **306**, 1146–1149.

Amrhein, V. & Greenland, S. (2017) Remove, rather than redefine, statistical significance. *Nature Human Behaviour*, 1.

Anderegg, W.R.L., Klein, T., Bartlett, M., Sack, L., Pellegrini, A.F.A., Choat, B., & Jansen, S. (2016) Meta-analysis reveals that hydraulic traits explain cross-species patterns of drought-induced tree mortality across the globe. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 5024–5029.

Andrew, N.R. & Hughes, L. (2005) Herbivore damage along a latitudinal gradient: relative impacts of different feeding guilds. *Oikos*, **108**, 176–182.

van Asch, M. & Visser, M.E. (2007) Phenology of forest caterpillars and their host trees: the importance of synchrony. *Annual Review of Entomology*, **52**, 37–55.

Attwood, S.J., Maron, M., House, A.P.N., & Zammit, C. (2008) Do arthropod assemblages display globally consistent responses to intensified agricultural land use and management? *Global Ecology and Biogeography*, **17**, 585–599.

Baeten, L., Verheyen, K., Wirth, C., et al. (2013) A novel comparative research platform designed to determine the functional significance of tree species diversity in European forests. *Perspectives in Plant Ecology, Evolution and Systematics*, **15**, 281–291.

Baker, M. (2016) 1,500 scientists lift the lid on reproducibility. *Nature*, **533**, 452–454.

Bale, J.S., Masters, G.J., Hodkinson, I.D., Awmack, C., Bezemer, T.M., Brown, V.K., Butterfield, J., Buse, A., Coulson, J.C., & Farrar, J. (2002) Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biology*, **8**, 1–16.

Barbati, A., Corona, P., & Marchetti, M. (2007) *European forest types: Categories and types for sustainable forest management reporting and policy.* European Environment Agency.

Barbosa, P., Hines, J., Kaplan, I., Martinson, H., Szczepaniec, A., & Szendrei, Z. (2009) Associational resistance and associational susceptibility: having right or wrong neighbors. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 1–20.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., & Ferrer, E.A. (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, **471**, 51–57.

Barrufol, M., Schmid, B., Bruelheide, H., Chi, X., Hector, A., Ma, K., Michalski, S., Tang, Z., & Niklaus, P.A. (2013) Biodiversity promotes tree growth during succession in subtropical forest. *PLOS ONE*, **8**, e81246.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.

Beech, E., Rivers, M., Oldfield, S., & Smith, P.P. (2017) GlobalTreeSearch: The first complete global database of tree species and country distributions. *Journal of Sustainable Forestry*, **36**, 454–489.

Benítez-López, A., Alkemade, R., & Verweij, P.A. (2010) The impacts of roads and other infrastructure on mammal and bird populations: A meta-analysis. *Biological Conservation*, **143**, 1307–1316.

Benjamin, D.J., Berger, J.O., Johannesson, M., *et al.* (2017) Redefine statistical significance. *Nature Human Behaviour*, 1.

Bernays, E.A., Bright, K.L., Gonzalez, N., & Angel, J. (1994) Dietary mixing in a generalist herbivore: tests of two hypotheses. *Ecology*, **75**, 1997–2006.

Blanco, V., Brown, C., & Rounsevell, M. (2015) Characterising forest owners through their objectives, attributes and management strategies. *European Journal of Forest Research*, **134**, 1027–1041.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2011) *Introduction to meta-analysis.* John Wiley & Sons, Ltd,

Boyd, I.L., Freer-Smith, P.H., Gilligan, C.A., & Godfray, H.C.J. (2013) The consequence of tree pests and diseases for ecosystem services. *Science*, **342**, 1235773.

Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5–32.

Brockerhoff, E.G., Barbaro, L., Castagneyrol, B., Forrester, D.I., Gardiner, B., González-Olabarria, J.R., Lyver, P.O., Meurisse, N., Oxbrough, A., Taki, H., Thompson, I.D., van der Plas, F., & Jactel, H. (2017) Forest biodiversity, ecosystem functioning and the provision of ecosystem services. *Biodiversity and Conservation*, **26**, 3005–3035.

Brown, B.J. & Ewel, J.J. (1987) Herbivory in complex and simple tropical successional ecosystems. *Ecology*, **68**, 108–116.

Bruelheide, H., Böhnke, M., Both, S., *et al.* (2011) Community assembly during secondary forest succession in a Chinese subtropical forest. *Ecological Monographs*, **81**, 25–41.

Bruelheide, H., Nadrowski, K., Assmann, T., *et al*. (2014) Designing forest biodiversity experiments: general considerations illustrated by a new large experiment in subtropical China. *Methods in Ecology and Evolution*, **5**, 74–89.

Bukoski, J.J., Broadhead, J.S., Donato, D.C., Murdiyarso, D., & Gregoire, T.G. (2017) The use of mixed effects models for obtaining low-cost ecosystem carbon stock estimates in mangroves of the Asia-Pacific. *PLOS ONE*, **12**, e0169096.

van Buuren, S. & Groothuis-Oudshoorn, K. (2011) mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, **45**, 1–67.

Byrom, A.E., Innes, J., & Binny, R.N. (2016) A review of biodiversity outcomes from possum-focused pest control in New Zealand. *Wildlife Research*, **43**, 228–253.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., & Naeem, S. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.

Cardinale, B.J., Palmer, M.A., & Collins, S.L. (2002) Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature*, **415**, 426–429.

Carlson, C.J., Burgio, K.R., Dougherty, E.R., Phillips, A.J., Bueno, V.M., Clements, C.F., Castaldo, G., Dallas, T.A., Cizauskas, C.A., Cumming, G.S., Doña, J., Harris, N.C., Jovani, R., Mironov, S., Muellerklein, O.C., Proctor, H.C., & Getz, W.M. (2017) Parasite biodiversity faces extinction and redistribution in a changing climate. *Science Advances*, **3**, e1602422.

Castagneyrol, B., Giffard, B., Pere, C., & Jactel, H. (2013) Plant apparency, an overlooked driver of associational resistance to insect herbivory. *Journal of Ecology*, **101**, 418–429.

Castagneyrol, B., Jactel, H., Vacher, C., Brockerhoff, E.G., & Koricheva, J. (2014) Effects of plant phylogenetic diversity on herbivory depend on herbivore specialization. *Journal of Applied Ecology*, **51**, 134–141.

CBD (2010) Global Biodiversity Outlook 3.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., & Palmer, T.M. (2015) Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, **1**, e1400253.

Ceballos, G., Ehrlich, P.R., & Dirzo, R. (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E6089–E6096.

Chalmers, T.C., Matta, R.J., Smith, H., Jr, & Kunzler, A.M. (1977) Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *New England Journal of Medicine*, **297**, 1091–1096.

Chamagne, J., Tanadini, M., Frank, D., Matula, R., Paine, C.E.T., Philipson, C.D., Svátek, M., Turnbull, L.A., Volařík, D., & Hector, A. (2017) Forest diversity promotes individual tree growth in central European forest stands. *Journal of Applied Ecology*, **54**, 71–79.

Champagne, E., Tremblay, J.-P., & Côté, S.D. (2016) Spatial extent of neighboring plants influences the strength of associational effects on mammal herbivory. *Ecosphere*, **7**, e01371.

Cheung, M.W.-L. (2014) metaSEM: an R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, **5**, 1521.

Cheung, M.W.-L. (2015) *Meta-Analysis: A Structural Equation Modeling Approach*. John Wiley & Sons.

Chisholm, R.A., Muller-Landau, H.C., Abdul Rahman, K., *et al.* (2013) Scale-dependent relationships between tree species richness and ecosystem function in forests. *Journal of Ecology*, **101**, 1214–1224.

Clissold, F.J., Coggan, N., & Simpson, S.J. (2013) Insect herbivores can choose microclimates to achieve nutritional homeostasis. J*ournal of Experimental Biology*, **216**, 2089–2096.

Coley, P.D. (1983) Herbivory and defensive characteristics of tree species in a lowland tropical forest. *Ecological Monographs*, **53**, 209–229.

Coley, P.D. & Barone, J.A. (1996) Herbivory and plant defenses in tropical forests. *Annual Review of Ecology and Systematics*, **27**, 305–335.

Costantini, D., Edwards, D.P., & Simons, M.J.P. (2016) Life after logging in tropical forests of Borneo: A meta-analysis. *Biological Conservation*, **196**, 182–188.

Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'neill, R.V., & Paruelo, J. (1997) The value of the world's ecosystem services and natural capital. *Nature*, **387**, 253–260.

Craven, D., Thakur, M.P., Cameron, E.K., Frelich, L.E., Beauséjour, R., Blair, R.B., Blossey, B., Burtis, J., Choi, A., Dávalos, A., & Others (2017) The unseen invaders: introduced earthworms as drivers of change in plant communities in North American forests (a meta-analysis). *Global Change Biology*, **23**, 1065–1074.

Crouzeilles, R., Curran, M., Ferreira, M.S., Lindenmayer, D.B., Grelle, C.E.V., & Rey Benayas, J.M. (2016) A global meta-analysis on the ecological drivers of forest restoration success. *Nature Communications*, **7**, 11666.

Crowther, T.W., Glick, H.B., Covey, K.R., *et al.* (2015) Mapping tree density at a global scale. *Nature*, **525**, 201–205.

David Ford, E. (2000) *Scientific Method for Ecological Research.* Cambridge University Press.

De Beenhouwer, M., Aerts, R., & Honnay, O. (2013) A global meta-analysis of the biodiversity and ecosystem service benefits of coffee and cacao agroforestry. *Agriculture, Ecosystems & Environment*, **175**, 1–7.

Del Toro, I., Ribbons, R.R., & Pelini, S.L. (2012) The little things that run the world revisited: a review of ant-mediated ecosystem services and disservices (Hymenoptera: Formicidae). *Myrmecological News*, **17**, 133–146.

Díaz, S., Demissew, S., Carabias, J., Joly, C., Lonsdale, M., Ash, N., Larigauderie, A., Adhikari, J.R., Arico, S., Báldi, A., & Others (2015) The IPBES conceptual framework—connecting nature and people. *Current Opinion in Environmental Sustainability*, **14**, 1–16.

Díaz, S., Fargione, J., Chapin, F.S.I., & Tilman, D. (2006) Biodiversity loss threatens human well-being. *PLOS Biology*, **4**, e277.

Dieler, J., Uhl, E., Biber, P., Müller, J., Rötzer, T., & Pretzsch, H. (2017) Effect of forest stand management on species composition, structural diversity, and productivity in the temperate zone of Europe. *European Journal of Forest Research*, **136**, 739–766.

Dove, N.C. & Hart, S.C. (2017) Fire reduces fungal species richness and in situ mycorrhizal colonization: a meta-analysis. *Fire Ecology*, **13**, 37–65.

Dowle, M., Srinivasan, A., Short, T., from R Saporta, S.L.W.C., & Antonyan, E. (2015) data.table: extension of Data.frame.

Duffy, J.E., Godwin, C.M., & Cardinale, B.J. (2017) Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature*, **549**, 261–264.

Duraiappah, A.K., Naeem, S., Agardy, T., & Assessment, M.E. (2005) *Ecosystems and human well-being: biodiversity synthesis.* Island Press Washington DC,

Dyer, L.A., Singer, M.S., Lill, J.T., Stireman, J.O., Gentry, G.L., Marquis, R.J., Ricklefs, R.E., Greeney, H.F., Wagner, D.L., Morais, H.C., Diniz, I.R., Kursar, T.A., & Coley, P.D. (2007) Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, **448**, 696–U9.

Eisenhauer, N., Reich, P.B., & Scheu, S. (2012) Increasing plant diversity effects on productivity with time due to delayed soil biota effects on plants. *Basic and Applied Ecology*, **13**, 571–578.

Ellington, E.H., Bastille-Rousseau, G., Austin, C., Landolt, K.N., Pond, B.A., Rees, E.E., Robar, N., & Murray, D.L. (2015) Using multiple imputation to estimate missing data in meta-regression. *Methods in Ecology and Evolution*, **6**, 153–163.

Erskine, P.D., Lamb, D., & Bristow, M. (2006) Tree species diversity and ecosystem function: Can tropical multi-species plantations generate greater productivity? *Forest Ecology and Management*, **233**, 205–210.

European Commision (2013) A new EU Forest Strategy: for forests and the forest-based sector.

Fichtner, A., Härdtle, Li, Y., Bruelheide, H., Kunz, M., & von Oheimb, G. (2017) From competition to facilitation: how tree species respond to neighbourhood diversity. *Ecology letters*, **7**, 892-900.

Fidler, F., Chee, Y.E., Wintle, B.C., Burgman, M.A., McCarthy, M.A., & Gordon, A. (2017) Metaresearch for evaluating reproducibility in ecology and evolution. *Bioscience*, **67**, 282–289.

Firn, J., Erskine, P.D., & Lamb, D. (2007) Woody species diversity influences productivity and soil nutrient availability in tropical plantations. *Oecologia*, **154**, 521–533.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., Renner, S., Schöning, I., Schumacher, U., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.-D., & Weisser, W.W. (2010) Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. *Basic and Applied Ecology*, **11**, 473–485.

Fisher, R.A. (1921) On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.

Fontaine, J.B. & Kennedy, P.L. (2012) Meta-analysis of avian and small-mammal response to fire severity and fire surrogate treatments in U.S. fire-prone forests. *Ecological Applications*, **22**, 1547–1561.

Forister, M.L., Novotny, V., Panorska, A.K., Baje, L., Basset, Y., Butterill, P.T., Cizek, L., Coley, P.D., Dem, F., & Diniz, I.R. (2015) The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 442–447.

Forrester, D.I. & Bauhus, J. (2016) A review of processes behind diversity-productivity relationships in forests. *Current Forestry Reports*, **2**, 45–61.

Forsman, J.T., Reunanen, P., Jokimäki, J., & Mönkkönen, M. (2010) The effects of small-scale disturbance on forest birds: a meta-analysis. *Canadian Journal of Forest Research*, **40**, 1833–1842.

Fowler, N. (1990) The 10 most common statistical errors. *Bulletin of the Ecological Society of America*, **71**, 161–164.

Freedman, L.P., Cockburn, I.M., & Simcoe, T.S. (2015) The economics of reproducibility in preclinical research. *PLOS Biology*, **13**, e1002165.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., & Ringvall, A.H. (2014) Adapting national forest inventories to changing requirements - the case of the Swedish national forest inventory at the turn of the 20th century. *Silva Fennica*, **48**, 1095.

Furukawa, T.A., Barbui, C., Cipriani, A., Brambilla, P., & Watanabe, N. (2006) Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology*, **59**, 7–10.

Fuzessy, L.F., Cornelissen, T.G., Janson, C., & Silveira, F.A.O. (2016) How do primates affect seed germination? A meta-analysis of gut passage effects on neotropical plants. *Oikos*, **125**, 1069–1080.

Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., Ruiz-Jaen, M.C., Fröberg, M., Stendahl, J., & Philipson, C.D. (2013) Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications*, **4**, 1340.

García-Palacios, P., Vandegehuchte, M.L., Shaw, E.A., Dam, M., Post, K.H., Ramirez, K.S., Sylvain, Z.A., Tomasel, C.M., & Wall, D.H. (2015) Are there links between responses of soil microbes and ecosystem functioning to elevated CO2, N deposition and warming? A global perspective. *Global Change Biology*, **21**, 1590–1600.

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 in Ecology and Evolution*, **8**, 777–784.

Gillman, L.N., Wright, S.D., Cusens, J., McBride, P.D., Malhi, Y., & Whittaker, R.J. (2015) Latitude, productivity and species richness. *Global Ecology and Biogeography*, **24**, 107–117.

Glass, G.V. (1976) Primary, secondary, and meta-Analysis of research. *Educational Researcher*, **5**, 3–8.

Gómez-Aparicio, L. & Lortie, C.J. (2014) Advancing plant ecology through meta-analyses. *Journal of Ecology*, **102**, 823–827.

Gram, W.K., Sork, V.L., Marquis, R.J., Renken, R.B., Clawson, R.L., Faaborg, J., Fantz, D.K., Le Corff, J., Lill, J., & Porneluzi, P.A. (2001) Evaluating the effects of ecosystem management: A case study in a Missouri Ozark forest. *Ecological Applications*, **11**, 1667–1679.

Gray, M.A., Baldauf, S.L., Mayhew, P.J., & Hill, J.K. (2007) The response of avian feeding guilds to tropical forest disturbance. *Conservation Biology*, **21**, 133–141.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., & Altman, D.G. (2016) Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, **31**, 337–350.

de Groot, R.S., Alkemade, R., Braat, L., Hein, L., & Willemen, L. (2010) Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. *Ecological Complexity*, **7**, 260–272.

Grossman, J.J., Vanhellemont, M., Barsoum, N., *et al.* (2018) Synthesis and future research directions linking tree diversity to growth, survival, and damage in a global network of tree diversity experiments. *Environmental and Experimental Botany*, (in press).

Guo, Q. & Ren, H. (2014) Productivity as related to diversity and age in planted versus natural forests. *Global Ecology and Biogeography*, **23**, 1461–1471.

Haase, J., Castagneyrol, B., Cornelissen, J.H.C., Ghazoul, J., Kattge, J., Koricheva, J., Scherer-Lorenzen, M., Morath, S., & Jactel, H. (2015) Contrasting effects of tree diversity on young tree growth and resistance to insect herbivores across three biodiversity experiments. *Oikos*, **124**, 1674–1685.

Hambäck, P.A., Inouye, B.D., Andersson, P., & Underwood, N. (2014) Effects of plant neighborhoods on plant–herbivore interactions: resource dilution and associational effects. *Ecology*, **95**, 1370–1383.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., & Townshend, J.R.G. (2013) High-resolution global maps of 21st-century forest cover change. *Science*, **342**, 850–853.

Hartig, G.L. (1791) *Anweisung zur Holzzucht für Förster.* Neue akademische Buchhandlung,

Harzing, A.W. (2007) *Publish or Perish.*

Hedges, L.V. & Olkin, I. (2014) *Statistical Methods for Meta-Analysis.* Academic Press.

Hevia, V., Martín-López, B., & Palomo, S. (2017) Trait-based approaches to analyze links between the drivers of change and ecosystem services: Synthesizing existing evidence and future challenges. *Ecology*, **7**, 831–844.

Hijmans, R.J. (2013) *raster: Geographic Data Analysis and Modeling.*

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hijmans, R.J., Cameron, S., Parra, J., Jones, P., Jarvis, A., & Richardson, K. (2008) *WorldClim.*

Hillebrand, H. (2004) On the generality of the latitudinal diversity gradient. *The American Naturalist*, **163**, 192–211.

Hillebrand, H. & Gurevitch, J. (2013) Reporting standards in experimental studies. *Ecology Letters*, **16**, 1419–1420.

Hill, J.K., Gray, M.A., Khen, C.V., Benedick, S., Tawatao, N., & Hamer, K.C. (2011) Ecological impacts of tropical forest fragmentation: how consistent are patterns in species richness and nestedness? *Philosophical Transactions of the Royal Society of London. Series B*, **366**, 3265–3276.

Ho, D.E., Imai, K., King, G., & Stuart, E.A. (2007) Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, **15**, 199–236.

Ho, D., Imai, K., King, G., & Stuart, E. (2011) MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, **42**, 1–28.

Homeier, J., Breckle, S.-W., Günter, S., Rollenbeck, R.T., & Leuschner, C. (2010) Tree diversity, forest structure and productivity along altitudinal and topographical gradients in a species-rich Ecuadorian montane rain forest. *Biotropica*, **42**, 140–148.

Hooper, D.U., Adair, E.C., Cardinale, B.J., Byrnes, J.E.K., Hungate, B.A., Matulich, K.L., Gonzalez, A., Duffy, J.E., Gamfeldt, L., & O'Connor, M.I. (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, **486**, 105–108.

Hooper, D.U., Chapin, F.S., Iii, Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., & Naeem, S. (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, **75**, 3–35.

Hublin, J.-J., Ben-Ncer, A., Bailey, S.E., Freidline, S.E., Neubauer, S., Skinner, M.M., Bergmann, I., Le Cabec, A., Benazzi, S., Harvati, K., & Gunz, P. (2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature*, **546**, nature22336.

Idris, N.R.N. (2011) Estimating the bias in meta-analysis estimates for continuous data with non-random missing study variance. *Matematika*, **27**, 121–128.

Idris, N.R.N., Abdullah, M.H., & Tolos, S.M. (2013) Performance of selected imputation techniques for missing variances in meta-analysis. *Journal of Physics. Conference Series*, **435**, 012037.

Idris, N.R.N. & Robertson, C. (2009) The effects of imputing the missing standard deviations on the standard error of meta-analysis estimates. *Communications in Statistics - Simulation and Computation*, **38**, 513–526.

Isbell, F., Craven, D., Connolly, J., *et al.* (2015) Biodiversity increases the resistance of ecosystem productivity to climate extremes. *Nature*, **526**, 574–577.

IUCN (2012) Available at: https://www.iucn.org/content/facts-and-figures-forests.

Jacob, M., Leuschner, C., & Thomas, F.M. (2010) Productivity of temperate broad-leaved forest stands differing in tree species diversity. *Annals of Forest Science*, **67**, 503.

Jactel, H., Birgersson, G., Andersson, S., & Schlyter, F. (2011) Non-host volatiles mediate associational resistance to the pine processionary moth. *Oecologia*, **166**, 703–711.

Jactel, H. & Brockerhoff, E.G. (2007) Tree diversity reduces herbivory by forest insects. *Ecology Letters*, **10**, 835–848.

Jarvinen, A. (1991) A meta-analytic study of the effects of female age on laying-date and clutch-size in the Great Tit *Parus major* and the Pied Flycatcher *Ficedula hypoleuca*. *The Ibis*, **133**, 62–67.

Jeffries, J.M., Marquis, R.J., & Forkner, R.E. (2006) Forest age influences oak insect herbivore community structure, richness, and density. *Ecological Applications*, **16**, 901–912.

Jolani, S., Debray, T.P.A., Koffijberg, H., van Buuren, S., & Moons, K.G.M. (2015) Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Statistics in Medicine*, **34**, 1841–1863.

Jucker, T., Avăcăriței, D., Bărnoaiea, I., Duduman, G., Bouriaud, O., & Coomes, D.A. (2016) Climate modulates the effects of tree diversity on forest productivity. *Journal of Ecology*, **104**, 388–398.

Jucker, T., Bouriaud, O., Avacaritei, D., & Coomes, D.A. (2014a) Stabilizing effects of diversity on aboveground wood production in forest ecosystems: linking patterns and processes. *Ecology Letters*, **17**, 1560–1569.

Jucker, T., Bouriaud, O., Avacaritei, D., Dănilă, I., Duduman, G., Valladares, F., & Coomes, D.A. (2014b) Competition for light and water play contrasting roles in driving diversity–productivity relationships in Iberian forests. *Journal of Ecology*, **102**, 1202–1213.

Keenan, R.J., Reams, G.A., Achard, F., de Freitas, J.V., Grainger, A., & Lindquist, E. (2015) Dynamics of global forest area: results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, **352**, 9–20.

Kim, T.N. & Underwood, N. (2015) Plant neighborhood effects on herbivory: damage is both density and frequency dependent. *Ecology*, **96**, 1431–1437.

Knops, J.M.H., Tilman, D., Haddad, N.M., Naeem, S., Mitchell, C.E., Haarstad, J., Ritchie, M.E., Howe, K.M., Reich, P.B., Siemann, E., & Groth, J. (1999) Effects of plant species richness on invasion dynamics, disease outbreaks, insect abundances and diversity. *Ecology Letters*, **2**, 286–293.

Koellner, T. & Scholz, R.W. (2008) Assessment of land use impacts on the natural environment. *International Journal of Life Cycle Assessment*, **13**, 32.

Koricheva, J. (2002) Mixed forest stands as a means of sustainable forest management: project presentation and first results. *Proceedings of the SUFOR International Workshop*, 152–153.

Koricheva, J., Gurevitch, J., & Mengersen, K. (2013) *Handbook of Meta-Analysis in Ecology and Evolution.* Princeton University Press.

Kozlov, M.V. (2008) Losses of birch foliage due to insect herbivory along geographical gradients in Europe: a climate-driven pattern? *Climatic Change*, **87**, 107–117.

Kozlov, M.V., Lanta, V., Zverev, V., & Zvereva, E.L. (2015) Global patterns in background losses of woody plant foliage to insects. *Global Ecology and Biogeography*, **24**, 1126–1135.

Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5925–5930.

Kurtto, A., Sennikov, A., Lampinen, R., & Others (2013) *Atlas Florae Europaeae.* The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo.

Larsen, B.B., Miller, E.C., Rhodes, M.K., & Wiens, J.J. (2017) Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *The Quarterly Review of Biology*, **92**, 229–265.

Lasky, J.R., Uriarte, M., Boukili, V.K., Erickson, D.L., John Kress, W., & Chazdon, R.L. (2014) The relationship between tree biodiversity and biomass dynamics changes with tropical forest succession. *Ecology Letters*, **17**, 1158–1167.

Latham, J., Cumani, R., Rosati, I., & Bloise, M. (2014) Global land cover share (GLC-SHARE) database beta-release version 1.0-2014. *FAO: Rome, Italy*.

Lee, C.J. & Moher, D. (2017) Promote scientific integrity via journal peer review data. *Science*, **357**, 256–257.

Lemoine, N.P., Burkepile, D.E., & Parker, J.D. (2014) Variable effects of temperature on insect herbivory. *PeerJ*, **2**, e376.

Lepš, J. (2004) What do the biodiversity experiments tell us about consequences of plant species loss in the real world? *Basic and Applied Ecology*, **5**, 529–534.

Lewandowska, A.M., Biermann, A., Borer, E.T., *et al.* (2016) The influence of balanced and imbalanced resource supply on biodiversity-functioning relationship across ecosystems. *Philosophical Transactions of the Royal Society B*, **371**.

Liang, J., Crowther, T.W., Picard, N., *et al.* (2016) Positive biodiversity-productivity relationship predominant in global forests. *Science*, **354**, aaf8957.

Liang, J., Zhou, M., Tobin, P.C., McGuire, A.D., & Reich, P.B. (2015) Biodiversity influences plant productivity through niche-efficiency. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 5738–5743.

Lindsay, D.S. (2015) Replication in psychological science. *Psychological Science*, **26**, 1827–1832.

Li, T., Puhan, M.A., Vedula, S.S., Singh, S., Dickersin, K., & Ad Hoc Network Meta-Analysis Methods Meeting Working Group (2011) Network meta-analysis-highly attractive but more methodological research is needed. *BMC medicine*, **9**, 79.

Long, J.N. & Shaw, J.D. (2010) The influence of compositional and structural diversity on forest productivity. *Forestry*, **83**, 121–128.

Loreau, M. & Hector, A. (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature*, **412**, 72–76.

Madrigal-González, J., Ruiz-Benito, P., Ratcliffe, S., Calatayud, J., Kändler, G., Lehtonen, A., Dahlgren, J., Wirth, C., & Zavala, M.A. (2016) Complementarity effects on tree growth are contingent on tree size and climatic conditions across Europe. *Scientific Reports*, **6**, 32233.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2015) cluster: cluster analysis basics and extensions.

Magura, T., Lövei, G.L., & Tóthmérész, B. (2017) Edge responses are different in edges under natural versus anthropogenic influence: a meta-analysis using ground beetles. *Ecology and Evolution*, **7**, 1009–1017.

Mäkipää, R. & Heikkinen, J. (2003) Large-scale changes in abundance of terricolous bryophytes and macrolichens in Finland. *Journal of Vegetation Science*, **14**, 497–508.

Marchant, L.F. & McGrew, W.C. (1991) Laterality of function in apes: a meta-analysis of methods. *Journal of Human Evolution*, **21**, 425–438.

Marczak, L.B., Thompson, R.M., & Richardson, J.S. (2007) Meta-analysis: trophic level, habitat, and productivity shape the food web effects of resource subsidies. *Ecology*, **88**, 140–148.

Marquis, R.J., Ricklefs, R.E., & Abdala-Roberts, L. (2012) Testing the low latitude/high defense hypothesis for broad-leaved tree species. *Oecologia*, **169**, 811–820.

Martin, G.N. & Clarke, R.M. (2017) Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, **8**, 523.

Martin, P.A., Newton, A.C., & Bullock, J.M. (2017) Impacts of invasive plants on carbon pools depend on both species’ traits and local climate. *Ecology*, **98**, 1026–1035.

McCrackin, M.L., Jones, H.P., Jones, P.C., & Moreno-Mateos, D. (2017) Recovery of lakes and coastal marine ecosystems from eutrophication: A global meta-analysis. *Limnology and Oceanography*, **62**, 507–518.

Meyer, S.T., Ebeling, A., Eisenhauer, N., *et al.* (2016) Effects of biodiversity strengthen over time as ecosystem functioning declines at low and increases at high biodiversity. *Ecosphere*, **7**, e01619.

Milligan, H.T. & Koricheva, J. (2013) Effects of tree species richness and composition on moose winter browsing damage and foraging selectivity: an experimental study. *Journal of Animal Ecology*, **82**, 739–748.

Mina, M., Huber, M.O., Forrester, D.I., Thürig, E., & Rohner, B. (2017) Multiple factors modulate tree growth complementarity in Central European mixed forests. *Journal of Ecology*, early view.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & PRISMA Group (2010) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery*, **8**, 336–341.

Moles, A.T., Bonser, S.P., Poore, A.G.B., Wallis, I.R., & Foley, W.J. (2011) Assessing the evidence for latitudinal gradients in plant defence and herbivory. *Functional Ecology*, **25**, 380–388.

Morey, R.D., Chambers, C.D., Etchells, P.J., Harris, C.R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C.C., Newman, D.P., Schönbrodt, F.D., Vanpaemel, W., Wagenmakers, E.-J., & Zwaan, R.A. (2016) The Peer Reviewers’ Openness Initiative: incentivizing open research practices through peer review. *Royal Society Open Science*, **3**, 150547.

Moser, W.K. & Hansen, M. (2009) The relationship between diversity and productivity in selected forests of the Lake States Region (USA): relative impact of species versus structural diversity. *Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium* (ed. by R.E. McRoberts, G.A. Reams, P.C. Van Deusen, and W.H. McWilliams), pp. 149–157. U.S. Department of Agriculture, Forest Service, Washington DC.

Myers, J.A. & Harms, K.E. (2009) Seed arrival, ecological filters, and plant species richness: a meta-analysis. *Ecology Letters*, **12**, 1250–1260.

Nadrowski, K., Wirth, C., & Scherer-Lorenzen, M. (2010) Is forest diversity driving ecosystem function and service? *Current Opinion in Environmental Sustainability*, **2**, 75–79.

Naeem, S., Duffy, J.E., & Zavaleta, E. (2012) The functions of biological diversity in an age of extinction. *Science*, **336**, 1401–1406.

Naeem, S., Thompson, L.J., Lawler, S.P., Lawton, J.H., Woodfin, R.M., & Others (1994) Declining biodiversity can alter the performance of ecosystems. *Nature*, **368**, 734–737.

Naik, G. (2017) Peer-review activists push psychology journals towards open data. *Nature*, **543**, 161.

Nakagawa, S. (2015) Missing data: mechanisms, methods, and messages. *Ecological Statistics: Contemporary Theory and Application* (ed. by G.A. Fox, S. Negrete-Yankelevich, and V.J. Sosa), pp. 81–105. OUP Oxford University Press.

Nakagawa, S. & Freckleton, R.P. (2011) Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behavioral Ecology and Sociobiology*, **65**, 103–116.

Nakagawa, S. & Hauber, M.E. (2011) Great challenges with few subjects: statistical strategies for neuroscientists. *Neuroscience and Biobehavioral Reviews*, **35**, 462–473.

Nakagawa, S. & Parker, T.H. (2015) Replicating research in ecology and evolution: feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, **13**, 88.

Nekola, J.C. & McGill, B.J. (2014) Scale dependency in the functional form of the distance decay relationship. *Ecography*, **37**, 309–320.

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

Nguyen, H., Herbohn, J., Firn, J., & Lamb, D. (2012) Biodiversity-productivity relationships in small-scale mixed-species plantations using native species in Leyte province, Philippines. *Forest Ecology and Management*, **274**, 81–90.

Novotny, V., Basset, Y., Miller, S.E., Weiblen, G.D., Bremer, B., Cizek, L., & Drozd, P. (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature*, **416**, 841–844.

Nuzzo, R. (2014) Scientific method: statistical errors. *Nature*, **506**, 150–152.

Ohse, B., Seele, C., Holzwarth, F., & Wirth, C. (2017) Different facets of tree sapling diversity influence browsing intensity by deer dependent on spatial scale. *Ecology and Evolution*, **7**, 6779–6789.

Ojea, E., Paulo A L, & Loureiro, M.L. (2010) Mapping biodiversity indicators and assessing biodiversity values in global forests. *Environmental & Resource Economics*, **47**, 329–347.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., & Wagner, H. (2011) *vegan: community ecology package.*

Open Science Collaboration (2015) PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science*, **349**, aac4716.

Otway, S.J., Hector, A., & Lawton, J.H. (2005) Resource dilution effects on specialist insect herbivores in a grassland biodiversity experiment. *Journal of Animal Ecology*, **74**, 234–240.

Paquette, A. & Messier, C. (2011) The effect of biodiversity on tree productivity: from temperate to boreal forests. *Global Ecology and Biogeography*, **20**, 170–180.

Parker, T.H., Nakagawa, S., Gurevitch, J., & IIEE (Improving Inference in Evolutionary Biology and Ecology) workshop participants (2016) Promoting transparency in evolutionary biology and ecology. *Ecology Letters*, **19**, 726–728.

Pearse, I.S. & Hipp, A.L. (2012) Global patterns of leaf defenses in oak species. *Evolution*, **66**, 2272–2286.

Pearson, K. (1904) Report on certain enteric fever inoculation statistics. *British Medical Journal*, **2**, 1243–1246.

Pennings, S.C., Ho, C.-K., Salgado, C.S., WIeski, K., Davé, N., Kunza, A.E., & Wason, E.L. (2009) Latitudinal variation in herbivore pressure in Atlantic coast salt marshes. *Ecology*, **90**, 183–195.

Pereira, H.M., Leadley, P.W., Proença, V., *et al.* (2010) Scenarios for global biodiversity in the 21st Century. *Science*, **330**, 1496–1501.

Philpott, S.M., Arendt, W.J., Armbrecht, I., Bichier, P., Diestch, T.V., Gordon, C., Greenberg, R., Perfecto, I., Reynoso-Santos, R., Soto-Pinto, L., Tejeda-Cruz, C., Williams-Linera, G., Valenzuela, J., & Zolotoff, J.M. (2008) Biodiversity loss in Latin American coffee landscapes: review of the evidence on ants, birds, and trees. *Conservation Biology*, **22**, 1093–1105.

Piotto, D. (2008) A meta-analysis comparing tree growth in monocultures and mixed plantations. *Forest Ecology and Management*, **255**, 781–786.

van der Plas, F., Manning, P., Allan, E., *et al.* (2016) Jack-of-all-trades effects drive biodiversity-ecosystem multifunctionality relationships in European forests. *Nature Communications*, **7**, 11109.

Polley, H., Schmitz, F., Hennig, P., & Franz, K. (2009) Germany. *National Forest Inventories: Pathways for Common Reporting* (ed. by E. Tomppo, T. Gschwantner, M. Lawrence, and R.E. McRoberts), pp. 223–243. Springer.

Potvin, C. & Gotelli, N.J. (2008) Biodiversity enhances individual performance but does not affect survivorship in tropical trees. *Ecology Letters*, **11**, 217–223.

Pretzsch, H. (2005) Diversity and productivity in forests: evidence from long-term experimental plots. *Forest Diversity and Function* (ed. by D.M. Scherer-Lorenzen, P.D.C. Körner, and P.D.E.-D. Schulze), pp. 41–64. Springer Berlin Heidelberg.

Price, P.W., Bouton, C.E., Gross, P., McPheron, B.A., Thompson, J.N., & Weis, A.E. (1980) Interactions among three trophic levels: influence of plants on interactions between insect herbivores and natural enemies. *Annual Review of Ecology and Systematics*, **11**, 41–65.

Price, P.W., Fernandes, G.W., Lara, A.C.F., Brawn, J., Barrios, H., Wright, M.G., Ribeiro, S.P., & Rothcliff, N. (1998) Global patterns in local number of insect galling species. *Journal of Biogeography*, **25**, 581–591.

Quartagno, M. & Carpenter, J.R. (2016) Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Statistics in Medicine*, **35**, 2938–2954.

Rasband, W.S. (2015) *ImageJ.* Bethesda, Maryland.

Ratcliffe, S., Holzwarth, F., Nadrowski, K., Levick, S., & Wirth, C. (2015) Tree neighbourhood matters – Tree species composition drives diversity–productivity patterns in a near-natural beech forest. *Forest Ecology and Management*, **335**, 225–234.

Ratcliffe, S., Liebergesell, M., Ruiz-Benito, P., Madrigal González, J., Muñoz Castañeda, J.M., Kändler, G., Lehtonen, A., Dahlgren, J., Kattge, J., Peñuelas, J., Zavala, M.A., & Wirth, C. (2016) Modes of functional biodiversity control on tree productivity across the European continent. *Global Ecology and Biogeography*, **25**, 251–262.

Ratcliffe, S., Wirth, C., Jucker, T., *et al.* (2017) Biodiversity and ecosystem functioning relations in European forests depend on environmental context. *Ecology letters*, **20**, 1414-1426.

Raubenheimer, D. & Simpson, S.J. (1999) Integrating nutrition: a geometrical approach. *Proceedings of the 10th International Symposium on Insect-Plant Relationships* pp. 67–82. Springer Netherlands.

Ravenek, J.M., Bessler, H., Engels, C., Scherer-Lorenzen, M., Gessler, A., Gockele, A., De Luca, E., Temperton, V.M., Ebeling, A., Roscher, C., Schmid, B., Weisser, W.W., Wirth, C., de Kroon, H., Weigelt, A., & Mommer, L. (2014) Long-term study of root biomass in a biodiversity experiment reveals shifts in diversity effects over time. *Oikos*, **123**, 1528–1536.

R Core Team (2013) *R: A Language and Environment for Statistical Computing.* Vienna, Austria.

Reich, P.B., Tilman, D., Isbell, F., Mueller, K., Hobbie, S.E., Flynn, D.F.B., & Eisenhauer, N. (2012) Impacts of biodiversity loss escalate through time as redundancy fades. *Science*, **336**, 589–592.

Reid, W.V., Mooney, H.A., Cropper, A., Capistrano, D., Carpenter, S.R., Chopra, K., Dasgupta, P., Dietz, T., Duraiappah, A.K., & Hassan, R. (2005) Ecosystems and human well-being: synthesis. *Millennium Ecosystem Assessment*.

Resche-Rigon, M., White, I.R., Bartlett, J.W., Peters, S.A.E., Thompson, S.G., & PROG-IMT Study Group (2013) Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine*, **32**, 4890–4905.

Rhainds, M. & English-Loeb, G. (2003) Testing the resource concentration hypothesis with tarnished plant bug on strawberry: density of hosts and patch size influence the interaction between abundance of nymphs and incidence of damage. *Ecological Entomology*, **28**, 348–358.

Riffell, S., Verschuyl, J., Miller, D., & Wigley, T.B. (2011) Biofuel harvests, coarse woody debris, and biodiversity – A meta-analysis. *Forest Ecology and Management*, **261**, 878–887.

del Río, M., Pretzsch, H., Ruíz-Peinado, R., *et al.* (2017) Species interactions increase the temporal stability of community productivity in *Pinus sylvestris-Fagus sylvatica* mixtures across Europe. *Journal of Ecology*, **105**, 1032–1043.

Robertson, C., Idris, N.R.N., & Boyle, P. (2004) Beyond classical meta-analysis: can inadequately reported studies be included? *Drug Discovery Today*, **9**, 924–931.

Rodríguez-Castañeda, G. (2013) The world and its shades of green: a meta-analysis on trophic cascades across temperature and precipitation gradients. *Global Ecology and Biogeography*, **22**, 118–130.

Rondeux, J. & Wagner, M. (2009) Belgium (Walloon Region). *National Forest Inventories: Pathways for Common Reporting* (ed. by E. Tomppo, T. Gschwantner, M. Lawrence, and R.E. McRoberts), pp. 73–87. Springer.

Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin*, **86**, 638.

Rossetti, M.R., Tscharntke, T., Aguilar, R., & Batáry, P. (2017) Responses of insect herbivores and herbivory to habitat fragmentation: a hierarchical meta-analysis. *Ecology Letters*, **20**, 264–272.

Ruddiman, W.F. (2003) The anthropogenic greenhouse era began thousands of years ago. *Climatic Change*, **61**, 261–293.

Ruiz-Benito, P., Gómez-Aparicio, L., Paquette, A., Messier, C., Kattge, J., & Zavala, M.A. (2014) Diversity increases carbon storage and tree productivity in Spanish forests. *Global Ecology and Biogeography*, **23**, 311–322.

Ruiz-Benito, P., Ratcliffe, S., Jump, A.S., Gómez-Aparicio, L., Madrigal-González, J., Wirth, C., Kändler, G., Lehtonen, A., Dahlgren, J., Kattge, J., & Zavala, M.A. (2017) Functional diversity underlies demographic responses to environmental variation in European forests. *Global Ecology and Biogeography*, **26**, 128–141.

Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., & Richardson, W.S. (1996) Evidence based medicine: what it is and what it isn't. *BMJ*, **312**, 71–72.

Salazar, D. & Marquis, R.J. (2012) Herbivore pressure increases toward the equator. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 12616–12620.

Scherer-Lorenzen, M. (2014) The functional role of biodiversity in the context of global change. *Forests and Global Change* (ed. by DA. Coomes & WA. Simonson), Cambridge University Press.

Scherer-Lorenzen, M., Potvin, C., Koricheva, J., Schmid, B., Hector, A., Bornik, Z., Reynolds, G., & Schulze, E.-D. (2005) The design of experimental tree plantations for functional biodiversity research. *Forest Diversity and Function* (ed. by M. Scherer-Lorenzen, C. Körner, and E.-D. Schulze), pp. 347–376. Springer Berlin Heidelberg.

Schnitzer, S.A. & Carson, W.P. (2016) Would ecology fail the repeatability test? *Bioscience*, **66**, 98–99.

Schooler, J.W. (2014) Metascience could rescue the "replication crisis." *Nature News*, **515**, 9.

Schuldt, A., Baruffol, M., Boehnke, M., Bruelheide, H., Haerdtle, W., Lang, A.C., Nadrowski, K., von Oheimb, G., Voigt, W., Zhou, H., & Assmann, T. (2010) Tree diversity promotes insect herbivory in subtropical forests of south-east China. *Journal of Ecology*, **98**, 917–926.

Schuldt, A., Bruelheide, H., Haerdtle, W., Assmann, T., Li, Y., Ma, K., von Oheimb, G., & Zhang, J. (2015) Early positive effects of tree species richness on herbivory in a large-scale forest biodiversity experiment influence tree growth. *Journal of Ecology*, **103**, 563–571.

Schwappach, A.F. (1912) *Ertragstafeln der wichtigeren Holzarten in tabellarischer und graphischer Form.* J. Neumann.

Seibold, S., Bässler, C., Brandl, R., Gossner, M.M., Thorn, S., Ulyshen, M.D., & Müller, J. (2015) Experimental studies of dead-wood biodiversity—a review identifying global gaps in knowledge. *Biological Conservation*, **191**, 139–149.

Shantz, A.A., Lemoine, N.P., & Burkepile, D.E. (2016) Nutrient loading alters the performance of key nutrient exchange mutualisms. *Ecology Letters*, **19**, 20–28.

Spake, R., Ezard, T.H.G., Martin, P.A., Newton, A.C., & Doncaster, C.P. (2015) A meta-analysis of functional group responses to forest recovery outside of the tropics. *Conservation Biology*, **29**, 1695–1703.

ter Steege, H., Pitman, N.C.A., Killeen, T.J., *et al.* (2015) Estimating the global conservation status of more than 15,000 Amazonian tree species. *Science Advances*, **1**, e1500936.

Sterne, J.A.C. & Egger, M. (2005) Regression methods to detect publication and other bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 99–110.

Stoepler, T.M. & Rehill, B. (2012) Forest habitat, not leaf phenotype, predicts late-season folivory of *Quercus alba* saplings. *Functional Ecology*, **26**, 1205–1213.

Su, Y.-S., Gelman, A., Hill, J., & Yajima, M. (2011) Multiple imputation with diagnostics (mi) in R: opening windows into the black box. *Journal of Statistical Software*, **45**, 1–31.

Szwagrzyk, J. & Gazda, A. (2007) Above-ground standing biomass and tree species diversity in natural stands of Central Europe. *Journal of Vegetation Science*, **18**, 555–562.

Tahvanainen, J.O. & Root, R.B. (1972) The influence of vegetational diversity on the population ecology of a specialized herbivore, *Phyllotreta cruciferae* (Coleoptera: Chrysomelidae). *Oecologia*, **10**, 321–346.

Thiessen Philbrook, H., Barrowman, N., & Garg, A.X. (2007) Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *Journal of Clinical Epidemiology*, **60**, 228–240.

Tilman, D. & Downing, J.A. (1994) Biodiversity and stability in grasslands. *Nature*, **367**, 363–365.

Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., & Siemann, E. (1997) The Influence of functional diversity and composition on ecosystem processes. *Science*, **277**, 1300–1302.

Tobner, C.M., Paquette, A., Gravel, D., Reich, P.B., Williams, L.J., & Messier, C. (2016) Functional identity is the main driver of diversity effects in young tree communities. *Ecology Letters*, **19**, 638–647.

Tobner, C.M., Paquette, A., Reich, P.B., Gravel, D., & Messier, C. (2014) Advancing biodiversity–ecosystem functioning science using high-density tree-based experiments over functional diversity gradients. *Oecologia*, **174**, 609–621.

Tomppo, E. & Tuomainen, T. (2009) Finland. *National Forest Inventories: Pathways for Common Reporting* (ed. by E. Tomppo, T. Gschwantner, M. Lawrence, and R.E. McRoberts), pp. 185–206. Springer.

Tonin, F.S., Rotta, I., Mendes, A.M., & Pontarolo, R. (2017) Network meta-analysis: a technique to gather evidence from direct and indirect comparisons. *Pharmacy Practice*, **15**, 943.

Trivedi, P., Delgado-Baquerizo, M., Anderson, I.C., & Singh, B.K. (2016) Response of Soil properties and microbial communities to agriculture: implications for primary productivity and soil health indicators. *Frontiers in Plant Science*, **7**, 990.

Trogisch, S., Schuldt, A., Bauhus, J., *et al.* (2017) Toward a methodical framework for comprehensively assessing forest multifunctionality. *Ecology and Evolution*, **7**, 10652–10674.

Turnbull, L.A., Isbell, F., Purves, D.W., Loreau, M., & Hector, A. (2016) Understanding the value of plant diversity for ecosystem functioning through niche theory. *Proceedings of the Royal Society B*, **283**.

Underwood, N., Inouye, B.D., & Hambäck, P.A. (2014) A conceptual framework for associational effects: when do neighbors matter and how would we know? *The Quarterly Review of Biology*, **89**, 1–19.

Unsicker, S.B., Oswald, A., Köhler, G., & Weisser, W.W. (2008) Complementarity effects through dietary mixing enhance the performance of a generalist insect herbivore. *Oecologia*, **156**, 313–324.

del-Val, E. & Armesto, J.J. (2010) Seedling mortality and herbivory damage in subtropical and temperate populations: testing the hypothesis of higher herbivore pressure toward the tropics. *Biotropica*, **42**, 174–179.

Van Buuren, S. (2012) *Flexible imputation of missing data.* CRC press.

Vayreda, J., Gracia, M., Canadell, J.G., & Retana, J. (2012) Spatial patterns and predictors of forest carbon stocks in Western Mediterranean. *Ecosystems*, **15**, 1258–1270.

Vellend, M., Baeten, L., Myers-Smith, I.H., Elmendorf, S.C., Beauséjour, R., Brown, C.D., De Frenne, P., Verheyen, K., & Wipf, S. (2013) Global meta-analysis reveals no net change in local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 19456–19459.

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S.W., & Garant, D. (2014) Drawing ecological inferences from coincident patterns of population-and community-level biodiversity. *Molecular Ecology*, **23**, 2890–2901.

Verheyen, K., Vanhellemont, M., Auge, H., *et al.* (2016) Contributions of a global network of tree diversity experiments to sustainable forest plantations. *Ambio*, **45**, 29–41.

Verschut, T.A., Becher, P.G., Anderson, P., & Hambäck, P.A. (2016) Disentangling associational effects: both resource density and resource frequency affect search behaviour in complex environments. *Functional Ecology*, *30*, 1826–1833.

Viechtbauer, W. (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, **30**, 1–48.

Vilà, M., Carrillo-Gavilán, A., Vayreda, J., Bugmann, H., Fridman, J., Grodzki, W., Haase, J., Kunstler, G., Schelhaas, M., & Trasobares, A. (2013) Disentangling biodiversity and climatic determinants of wood production. *PLOS ONE*, **8**, e53530.

Vilà, M., Vayreda, J., Comas, L., Ibáñez, J.J., Mata, T., & Obón, B. (2007) Species richness and wood production: a positive association in Mediterranean forests. *Ecology Letters*, **10**, 241–250.

Villaescusa, R. & Díaz, R. (1998) Segundo inventario forestal nacional (1986-1996). *Ministerio de Medio Ambiente, ICONA, Madrid*.

Villanueva, J.A. (2005) Tercer inventario forestal nacional (1997-2007). *Ministerio de Medio Ambiente, Madrid*.

Wang, J., Cheng, Y., Zhang, C., Zhao, Y., Zhao, X., & Von Gadow, K. (2016) Relationships between tree biomass productivity and local species diversity. *Ecosphere*, **7**, e01562.

Wasserstein, R.L. & Lazar, N.A. (2016) The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, **70**, 129–133.

Waters, C.N., Zalasiewicz, J., Summerhayes, C., *et al.* (2016) The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science*, **351**, aad2622.

Westgate, M.J., Tulloch, A.I.T., Barton, P.S., Pierson, J.C., & Lindenmayer, D.B. (2017) Optimal taxonomic groups for biodiversity assessment: a meta-analytic approach. *Ecography*, **40**, 539–548.

White, J.A. & Whitham, T.G. (2000) Associational susceptibility of cottonwood to a box elder herbivore. *Ecology*, **81**, 1795–1803.

Wickham, H. (2009) *ggplot2: elegant graphics for data analysis.* Springer New York.

Wickham, H. & Francois, R. (2015) dplyr: a grammar of data manipulation.

Wickham, H. & Others (2007) Reshaping data with the reshape package. *Journal of Statistical Software*, **21**, 1–20.

Wiebe, N., Vandermeer, B., Platt, R.W., Klassen, T.P., Moher, D., & Barrowman, N.J. (2006) A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of Clinical Epidemiology*, **59**, 342–353.

Wiedemann, E. (1943) Der Vergleich der Massenleistung des Mischbestandes mit dem Reinbestand. *Allgemeine Forst-und Jagdzeitung*, **119**, 123–132.

Wilke, C.O. (2016) cowplot: streamlined plot theme and plot annotations for "ggplot2."

Williams, L.J., Paquette, A., Cavender-Bares, J., Messier, C., & Reich, P.B. (2017) Spatial complementarity in tree crowns explains overyielding in species mixtures. *Nature Ecology & Evolution*, **1**, 0063.

Willms, J., Bartuszevige, A., Schwilk, D.W., & Kennedy, P.L. (2017) The effects of thinning and burning on understory vegetation in North America: a meta-analysis. *Forest Ecology and management*, **392**, 184–194.

Wolowski, M., Ashman, T.-L., & Freitas, L. (2014) Meta-analysis of pollen limitation reveals the relevance of pollination generalization in the Atlantic forest of Brazil. *PLOS ONE*, **9**, e89498.

Zaitsev, A.S., Gongalsky, K.B., Malmström, A., Persson, T., & Bengtsson, J. (2016) Why are forest fires generally neglected in soil fauna research? A mini-review. *Applied Soil Ecology*, **98**, 261–271.

Zhang, Y., Chen, H.Y.H., & Reich, P.B. (2012) Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *Journal of Ecology*, **100**, 742–749.

Zuur, A.F. & Ieno, E.N. (2016) A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, **7**, 636–645.

9.3 | List of Publications

van der Plas, F. *et al.* (2018) Continental mapping of forest ecosystem functions reveals a high but unrealized potential for forest multifunctionality. Ecology Letters, 21, 31-42.

Ratcliffe, S. *et al.* (2017) Biodiversity and ecosystem functioning relations in European forests depend on environmental context. *Ecology Letters*, **20**, 1414-1426.

Moreno-Mateos, *et al.* (2017) Will your paper be used in a meta-analysis? Make the reach of your research broader and longer lasting. *Methods in Ecology and Evolution*, **8**, 777-784.

Dittrich, A. *et al.* (2017) Mapping and analysing historical indicators of ecosystem services in Germany. *Ecological Indicators*, **75**, 101-110.

Kambach, S. *et al.* (2016) The impact of tree diversity on different aspects of insect herbivory along a global temperature gradient - A meta-analysis. *PLOS ONE*, **11**, e0165815.

Seppelt, R. *et al.* (2016) Harmonizing biodiversity conservation and productivity in the context of increasing demands on landscapes. *BioScience*, **66**, 890-896.

Kambach, S. *et al.* (2012) Human-induced disturbance alters pollinator communities in tropical mountain forests. *Diversity*, **5**, 1-4.

9.4 | Submitted Manuscripts

Kambach, S. *et al.* (re-submitted to *Ecography* with minor revisions) Of niches and distributions: range size increases with niche breadth both globally and regionally but not across scales for European vascular plants.

Kambach, S. *et al.* (in review for *Journal of Ecology*) How does tree species response to mixing compare between experimental and observational studies?

Kambach, S. *et al.* (in review for *Methods in Ecology and Evolution*) Consequences of multiple imputation of missing standard deviations in meta-analysis

Beckmann, M. *et al.* (submitted to *Nature Sustainability*) The simultaneous effects of land-use intensification on biodiversity and production: A global meta-analysis.

Baeten, L. *et al.* (submitted to *Journal of Applied Ecology*) Which tree species compositions maximize ecosystem functioning in European forest?

Sarathchandra *et al.* (in revision for *Diversity*) Significance of mangrove biodiversity conservation in fishery production and living conditions of coastal communities in Sri Lanka.

9.5 | Curriculum Vitae

02/2018 -	**Scientist** at the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany **Topic**: Beyond the growth-survival trade-off: A global analysis of demographic diversity and trade-offs in species-rich forests **PIs**: Dr. Nadja Rüger, Prof. Christian Wirth, Prof. Helge Bruelheide
07/2013 - 12/2017	**PhD studies** at the Helmholtz-Centre for Environmental Research, Germany, and the Martin-Luther-University, Halle, Germany **Topic**: Meta-analysis in forest biodiversity-ecosystem functioning research **Supervisors**: Prof. Helge Bruelheide, Prof. Ingolf Kühn, Prof. Ralf Seppelt
11/2012 - 06/2013	**Research stay** at the Université Picardie - Jules Verne, France, funded via the European Leonardo da Vinci-scholarship **Aim**: Learn and apply *R* to analyse range-size niche breadth relationships of European vascular plants **Supervisors**: Prof. Guillaume Decocq, Dr. Jonathan Lenoir
10/2006 - 07/2012	**Studies in Biology** at the Martin-Luther-University, Halle-Wittenberg, Germany **Thesis**: Differences in diversity and composition of pollinator guilds between montane forest and arrested succession in Bolivia **Supervisors**: Prof. Isabell Hensen, Dr. Matthias Schleuning
02/2010 - 04/2010	**Internship** in the Manu national park, Peru, as part of the university studies, together with the Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany **Task**: Fieldwork for Dr. Matthias Dehling

9.7 | Danksagung

Ich möchte mich bei all jenen Menschen, die mich in den letzten Jahren unterstützt, beraten, begleitet und motiviert haben; kurz: all jenen Menschen, dank welchen ich eine unvergessliche und (hoffentlich) erfolgreiche Promotionszeit hatte, bedanken.

Ich danke Prof. Helge Bruelheide für seine Unterstützung, Motivation, Vertrauen und die vielen Freiheiten, welcher er mir bei dieser Promotion und meinen weiteren Forschungsprojekten gab.

Prof. Ingolf Kühn möchte ich für seine Hilfe, Ratschläge und die Süßigkeiten in den AG-Sitzungen, zu welchen ich leider nicht sehr oft kommen konnte, danken.

Für die Chance, im Rahmen des SESYNC-LUBDES-Projektes Erfahrungen in einer internationalen Arbeitsgruppe und in Amerika sammeln zu können, möchte ich Prof. Ralf Seppelt herzlich danken.

Dr. Michael Beckmann möchte ich für die Betreuung und Unterstützung durch die ESCALATE Research School danken. Auch möchte ich der HIGRADE Research School und dem Helmholtz-Zentrum für Umweltforschung generell danken, welche mir meine Promotion, sowie viele interessante Weiterbildungen finanziert und ermöglicht haben.

I would like to thank all Co-authors fort their help and input, especially Dr. Bastien Castagneyrol and Prof. Jessica Gurevitch, who had major parts in developing and helping with the study 2 and 3 of this dissertation.

Tina Fuhrmann danke ich von Herzen für ihre Gelassenheit, ihren Optimismus, ihre Liebe, unsere zwei wundervollen Kinder und unsere gemeinsame Zukunft auf die ich mich freue.

Unseren Familien möchte ich für ihre Unterstützung und die Betreuung unserer Kinder danken, ohne welche mancher Forschungsaufenthalt nicht möglich gewesen wären.

Ich danke Sebastian Gebauer, Michael Köhler und allen Ehemaligen der Mittagsrunde für den guten Kuchen, Kaffee und unsere Gespräche.

9.6 | Eigenständigkeitserklärung

Hiermit erkläre ich, dass die vorliegende Arbeit mit dem Titel ˈMeta-analysis in Forest Biodiversity-Ecosystem Functioning Researchˈ bisher weder bei der Naturwissenschaftlichen Fakultät 1 der Martin-Luther-Universität Halle-Wittenberg noch bei einer anderen wissenschaftlichen Einrichtung zur Promotion vorgelegt wurde.

Weiterhin erkläre ich hiermit, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe und keine anderen, als die von mir angegebenen Quellen und Hilfsmittel genutzt habe. Wörtlich oder inhaltlich entnommene Stellen habe ich als solche kenntlich gemacht.

Ich erkläre außerdem, dass ich bisher keinen Promotionsversuch unternommen oder mich um einen Doktorgrad beworben habe.

Halle, den

Stephan Kambach