Effects of forestry practices on the conservation of soil and forest ecosystem health and functioning: an umbrella review protocol

Céline Moreaux[1]*, Philip A. Martin[1], Ana-Maria Hereş[1,2], Jorge Curiel Yuste[1]

[1] Basque Centre for Climate Change (BC3), Edificio sede no 1, planta 1, Parque científico UPV/EHU, Barrio Sarriena s/n, 48940, Leioa, Bizkaia, Spain

[2] Department of Forest Sciences, Transilvania University of Braşov, Sirul Beethoven -1, 500123 Braşov, Romania

* Corresponding author

Author email addresses: CM: celine.moreaux@bc3research.org; PAM: phil.martin.research@gmail.com; AMH: ana_heres@yahoo.com; JCY: jorge.curiel@bc3research.org

Abstract

Background: Forests host a majority of the world's terrestrial biodiversity and provide habitats and numerous resources, making them a vitally important ecosystem for plants, animals and humans alike. However, forest ecosystems are continuously lost under pressures such as land use change and fragmentation, logging activities and climate change. Timber and other forest resources are in high demand, so different economic and environmental interests lead to conflicts between resource extraction and biodiversity conservation. Nonetheless, forest management practices differ in their intensity and conclusively in their impacts on forest ecosystem health. Although various evidence syntheses have been conducted for specific forestry practices or geographic regions, a global overview of these impacts is still lacking. We use an umbrella review of existing evidence syntheses to fill this knowledge gap and to provide a basis for designing better informed forest management strategies and policies. We address the research questions: 1. How do different forestry practices affect the conservation of soil and forest ecosystem health and functioning? 2. Which evidence synthesis gaps exist with regards to the effects of forestry practices on forest and soil health? and 3. What is the quality and reliability of evidence from existing meta-analyses and syntheses on impacts of forestry practices on forest and soil health?

Methods: The review will be conducted using a standardised methodology for systematic reviews. We use relevant keywords on forest management and forest health to construct search strings for multiple published and grey literature databases. Inclusion criteria require that studies are meta-analyses reporting a quantitative effect size, focus on forest ecosystems, assess the effect of pre-defined forest management practices on pre-defined forest health indicators at different levels of intensity, and that the full manuscript is written in English. Relevant evidence syntheses are identified, collected and their quality critically appraised. A pre-defined list of data is extracted from the manuscripts into a global dataset, with a particular focus on the effect size and measures of heterogeneity. Finally, data synthesis will focus on narrative, tabular, and graphic summaries of the dataset to answer the research questions, highlight evidence gaps and make recommendations for future reviews, policies and decision-making.

Background

Forests cover around a third of the earth's land area, account for a majority of global gross primary production (Pan et al., 2013), and provide habitats for around two third of the world's terrestrial biodiversity. In addition, forests are important providers of ecosystem services on which people depend. They contribute to soil and watershed formation, air purification (Thompson et al., 2009), and provide a major carbon pool at a global scale, with more terrestrial carbon stored in forest biomass and soils than in the atmosphere (Pan et al., 2013). Thus, forests are vitally important in the attempts to mitigate and adapt to anthropogenic climate change mitigation as well as human well-being from local to global scales (Secretariat of the Convention on Biological Diversity, 2010).

However, forest ecosystems are increasingly under threat from land use change, fragmentation, diseases, natural and anthropogenic disturbances, as well as climate change (Hill et al., 2019). As a consequence, forests and forest biodiversity are being lost at a high rate and the capacity of forests to recover from these disturbances is jeopardised (Hansen et al., 2013; Ibáñez et al., 2019). Tree plantations, alongside afforestation and restoration, have been suggested as a way to mitigate losses of forest biodiversity and to provide some of the services associated with natural forests, such as carbon sequestration, fibre, timber and drinking water provision (Castaño-Villa et al., 2019; Wang et al., 2022). An additional benefit of forests managed for timber is that they are a major source of economic profits and provide employment and livelihoods to millions of people (Chaudhary et al., 2016).

However, rather than mitigating environmental problems, plantations often appear to negatively impact forest biodiversity and ecosystem functioning and can contribute to forest degradation at a large scale (Hua et al., 2022; Wang et al., 2022). The effects can vary in direction and intensity, and are highly dependent on the forest management practices employed as well as previous land use. Choice of tree species, intensity and frequency of harvests, species composition, and continuity of forest cover can all influence the effects of these practices on above- and belowground forest ecosystem health and resilience (see for instance Castaño-Villa et al., 2019; Clarke et al., 2021; Dieler et al., 2017; Duguid & Ashton, 2013; Gazol et al., 2021; Lázaro-Lobo et al., 2022).

Conclusively, conflicts between economic and environmental motivations influence forestry practices and can lead to trade-offs between different aspects of biodiversity and ecosystem functioning (Chaudhary et al., 2016). However, a comprehensive overview of these trade-offs between human exploitation of forest resources and their impacts on forest health is currently still lacking. Consequently, these trade-offs cannot be explicitly accounted for in forest policies and management strategies. Therefore, in this study we use an umbrella review methodology as an attempt to fill this evidence gap and to provide a knowledge basis for forest managers and decision makers.

Systematic and umbrella reviews

Motivated by the trade-offs described above, many primary studies have already outlined the effects of forest management practices on forest health and functioning (recent examples are for instance Hereş et al., 2021; Lázaro-Lobo et al., 2022; Leuschner et al., 2022; Naudts et al., 2016; Perring et al., 2018). These studies provide highly valuable insights into the possible consequences of management choices, but implementation of their results by practitioners and policy-makers is, in part, limited by increases in the volume of primary research and the scope of primary studies (Haddaway et al., 2018; Joanna Briggs Institute, 2014). In other words, the number of primary studies has increased so rapidly in recent decades that it is difficult to identify optimal practices for a

given case. This is complicated by the fact, that findings may be conflicting, depending on the geographical and research scope of the study, which makes it difficult for implementing bodies to make an informed choice. Furthermore, due to the design and logistics of primary studies, most of them focus solely on the impact of a single intervention on a small number of outcomes, meaning they are often very specific in their scope.

To address these issues, evidence syntheses have become a common tool to gather primary literature results and to conduct data analyses, in order to summarise information and identify generalities (for general guidelines see for instance Okoli, 2015; Pullin & Stewart, 2006). Evidence syntheses can furthermore be used to critically assess the methodological quality and reliability of primary studies, as well as to identify knowledge and syntheses gaps (Neal R. Haddaway & Pullin, 2014). Resulting assessments can be quantitative meta-analyses, using statistical methods to combine outcomes across data set, or qualitative, narrative syntheses, summarising results and drawing conclusions based on prose and experience (O'Leary et al., 2017). Within the environmental sciences, comprehensive guidelines and standards for evidence syntheses were for instance developed by the Collaboration for Environmental Evidence (CEE; Collaboration for Environmental Evidence, 2018).

However, as the number of evidence syntheses increases steadily – in line with increases in primary studies – practitioners again face a similar challenge to that seen for primary studies. To summarise this evidence and widen the scope of included results, another review level can be employed: *umbrella reviews*, also called reviews/synthesis/overview of reviews, or summaries of systematic reviews (Aromataris et al., 2015). These employ the same methodologies used for systematic reviews, with the sole difference that they focus only on the inclusion of studies that are narrative and/or quantitative syntheses. As these syntheses gather a large body of information available on a given topic, they make it possible to answer broad, complex questions, evaluate whether there are evidence gaps or redundancies and whether results are consistent (i.e. whether similar/contrasting results on a topic are found by multiple authors independently; Joanna Briggs Institute, 2014). This is crucial for the purpose of this study, which has the aims to identify general patterns in the effect of forest management practices on forest health across different meta-analyses and to detect research gaps that need to be addressed in greater detail in the future.

Finally, as with evidence syntheses of primary data, umbrella reviews can and should include a critical appraisal of the methodologies employed in the included studies and reflect on the consequences of methodological shortcomings and risks of biases within studies. Specific guidelines for umbrella reviews within environmental sciences do not exist yet, but for medicine and health sciences a manual was developed by the Joanna Briggs Institute (JBI), which we will adapt in the following for this review (Joanna Briggs Institute, 2014).

Objective of the review and research questions

Using an umbrella review, the objective of this study is to synthesise existing findings from meta-analyses on how forestry practices affect the conservation of soil and forest ecosystem health. Furthermore, the study serves to analyse whether this effect differs between more conservative (e.g. promotion of native species using long rotation ages) and less conservative practices (e.g. use of exotic species on short-rotation systems). Finally, it aims to identify gaps in research and to assess the quality of the conducted analyses using a standardised critical appraisal methodology.

Accordingly, the research questions of the study are as follows:

1. How do different forestry practices affect the conservation of soil and forest ecosystem health and functioning?
2. Which evidence synthesis gaps exist with regards to the effects of forestry practices on forest and soil health?
3. What is the quality and reliability of evidence from existing meta-analyses and syntheses on impacts of forestry practices on forest and soil health?

Question components

We defined six elements of the first research question to frame the scope of the umbrella review, namely *population, interventions/exposures, comparisons, outcomes* (the PICO elements; see Livoreil et al., 2017), as well as the *context/geographical scope* and *type of studies* included.

POPULATION: The review will include studies focused on forest ecosystems at above- or below-ground level. Forest definitions used in existing studies may differ, so we adhere to the respective definitions chosen by each publication.

INTERVENTIONS: Studies are included when they investigated the impact of *forest management practices* on forest ecosystems. These forestry interventions are limited to the choice of tree species (native, exotic), rotation length and stand age, forest structure and cover (even-, uneven-aged), forest type (broadleaved, conifer, mixed), as well as thinning and harvesting intensities (whole tree, stem-only, etc.).

COMPARISONS: We are interested in studies, which assessed forest management practices at varying levels of management intensity, such as harvesting intensity (clearcutting compared to selective cuttings), thinning intensity, species composition (monocultures compared to polycultures) and continuity of forest cover (even-aged compared to uneven aged forests).

OUTCOMES: We are interested in any outcomes related to forest health, defined at ecosystem, biodiversity and soil level as follows.

Forest ecosystem health: ecosystem processes that indicate the functioning of the forest ecosystem, using indicators on productivity, tree growth and mortality, the occurrence of tree predators and pests, forest natural regeneration capacity, forest resilience, recovery and resistance, as well as the intactness of the forest water and carbon cycles.

Forest biodiversity: richness and abundance of animal, plant and fungal species above and below-ground.

*Soil health***:** soil properties that inform about the physical, hydraulic, chemical and biological functioning of forest soils, using indicators on soil nutrients (C, N, P), soil biomass and organic matter, decomposition rates.

CONTEXT: The geographical scope of the study will be global. If possible, effect sizes and other data will be reported separately per climate zone, namely boreal, temperate, semi-arid, arid and tropical climates.

STUDY TYPES: For the review, only meta-analyses of multiple primary studies will be considered. These represent quantitative syntheses of primary study results on a selection of the intervention and outcome properties.

A detailed list of categories of all intervention and outcome indicators is given in Appendix I (Table A1 and Table A2).

In- and exclusion criteria

For a study to be eligible to be included into our review, it will have to fulfil the following **inclusion criteria**:

A. The study is a meta-analysis, systematic review or evidence synthesis of multiple primary studies that reports a quantitative effect size (*study type*).
B. The study is focused on forest ecosystems at above- or belowground level, including studies of forest soils (*population*).
C. The study assesses
    1. the effect of one or several of the defined forest management practices (*intervention*) on
    2. one or several of the defined forest health indicators (*outcome*).
D. The study must make a comparison between different levels of intensity of the selected forest management practices (e.g. having a baseline/control and treatment group). The study is also included, if it is a presence/absence comparison (e.g. managed/unmanaged forest) of a *specific* management practice (*comparison*).
E. The manuscript must be written in English (*language*).

During the screening process of titles and abstracts, articles will be retained for full-text screening if they fulfil the criteria A, B, and C. For the full-text review, articles have to fulfil all five criteria to be included.

Conversely, the following **exclusion criteria** apply:

a. The study is a qualitative (systematic) literature review that does not report findings through a quantitative effect size, or the study is solely a protocol for a systematic review or map.
b. The study is excluded, if it focuses on other land uses or ecosystems, such as grasslands, agricultural land, etc.
c. The study assesses effects of forestry practices or forest health indicators that are not outlined above in the PICO elements (e.g. grazing, controlled burning, fertilisation, machinery, post-disturbance or salvage logging, non-timber forest products/NTFP or non-wood forest products and services/NWFP&Ss), or does not specify individual forestry practices.
d. The study is excluded if it does not have a baseline, so it does not compare different intensity levels or presence/absence of a forest management practice, for instance if forest ecosystems are only included in comparison to other land uses or ecosystems, such as grasslands, agricultural land. Or the study compares different degrees of forest degradation (e.g. primary VS secondary forest, restored forest VS unrestored land, natural VS managed forest) without assessing the effect of individual forestry practices.
e. The full-text manuscript is written in a language other than English.

Further reasons for the exclusion of studies at full-text stage are:

f. The required data cannot be retrieved from the publication or after attempted contact with the author(s).

g. The data is a duplicate of another published dataset. In this case, the more comprehensive dataset is used for the analysis (Foo et al., 2021).

Review strategy

As indicated in the background section, umbrella reviews are a useful, but still rarely explored tool in ecology and climate change adaptation research. On the other hand, meta-analyses scrutinising the effects of human interventions on forest ecosystems are a much more frequently employed evidence tool (see for instance the list of benchmark studies below in Table 2). Given the broad range of data available from these analyses, we will use an umbrella review to synthesise and summarise large-scale effects of forest management practices on forest ecosystem health.

For this purpose, we adapt the methodology available from umbrella review guidelines from the health sciences for our purposes (Fusar-Poli & Radua, 2018; Joanna Briggs Institute, 2014). These are complemented by more specific guidelines for evidence syntheses within the environmental sciences, such as those by the Collaboration for Environmental Evidence (2020), Haddaway et al. (2018) and Livoreil et al. (2017). This includes the checklist and a flow diagram of the RepOrting standards for Systematic Evidence Syntheses (ROSES) (Appendix II; Table A3 and Figure A1), which will be filled out and attached for transparent documentation of the process (Haddaway et al., 2018). A ROSES checklist for systematic review protocols will be furthermore provided with this protocol to transparently document the preparatory steps taken.

Stake- and rightholder engagement

This umbrella review is part of a larger project on the development of soil conserving forestry practices and tools for the early detection of forest vulnerability, called ATLANTIS. This project consists of a range of analyses, including case studies and field experiments in Northern Spain.

The management and maintenance of forests is largely in the hands of forest technicians and managers, private land owners and government administrations. Thus, these actors do not only possess valuable knowledge, but also decision making and control capacities, which are pivotal for final choices on forest management plans. To be able to reflect as much of this insight as possible in the design of ATLANTIS, and in this umbrella review, a central goal of the project is to capture actors' preferences and existing areas of knowledge through direct engagement with them. This is in line with the recommendations of the CEE guidelines to involve actors into the evidence synthesis process, as they can play an important role in designing the research question and search procedure (Collaboration for Environmental Evidence, 2020). This engagement can help to prioritise research foci to ensure that the review output is considered relevant and endorsed by broader society. It can furthermore help to set the scope and goals of the project and to improve communication of the results. Stake- and rightholder engagement should take place at an early stage of the process, to identify preferences and needs of involved actors and collaborations in following project steps (N. R. Haddaway et al., 2017). In the ATLANTIS project, this was undertaken through the early dissemination of a survey on the management of Iberian Atlantic forests to a set of forestry associations, government entities, academic and innovation institutions, private forest owners and NGOs. The survey was designed so that actors can highlight particular interests in forest management practices and concerns about specific aspects of forest health. They were furthermore invited to participate in interviews and workshops, which are planned for a later stage of the project. In this process, we attempted to reach out to a balanced set of actors, to give space to different opinions, views and concerns. Stake- and rightholders were selected through purposive and

snowballing selection methods, so existing contacts were used to disseminate the survey and to ask for recommendations for additional recipients of the survey. As described by Haddaway et al. (2017), these selection methods can lead to identification and network biases, risking the exclusion of minorities and unknown groups. However, we attempted to mitigate these biases by using multiple starting points of contact with a very diverse range of backgrounds.

After the dissemination, the results from the survey were subsequently used to help identify research foci and search keywords for both forestry practices and forest ecosystem health indicators, by answering the following questions: (i) *Which are the forest management practices, actors have the most knowledge on?* (ii) *Which effects of forestry practices on forest ecosystem health are considered most pressing and relevant by actors? (iii) Which forestry practices do actors find most relevant for future research to improve forest health?*

Search engines

We will conduct the literature search of relevant peer-reviewed studies using the search engines Web of Science, SCOPUS and the CAB Forest Science Database. Multiple databases will be searched to ensure identifying more relevant peer-reviewed publications (Foo et al., 2021). In Web of Science, the search is conducted within the Core Collection based on ten indexes: *Journal Citation Indexes* (Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index), *Conference Proceedings* (Conference Proceedings Citation Index – Science as well as Social Science & Humanities), *Book Citation Index* (Book Citation Index-Science and Social Sciences & Humanities), and finally *Chemical Indexes* (Current Chemical Reactions, Index Chemicus).

Title, abstract and keywords will be included in the first screening stage of the search, full-text manuscripts in the second screening (see sub-heading *Screening stages* below). Furthermore, to include unpublished and grey literature, and to reduce publication bias, the first 300 results from Google Scholar will also be reviewed by the authors (Godin et al., 2015; Haddaway et al., 2015). Google Scholar is recognised as a valuable source for evidence reviews, if used as a complementary tool and not as a standalone search engine (Foo et al., 2021; Gusenbauer & Haddaway, 2020).

Search terms

Search terms were identified using a set of known relevant publications (see Table 2 below), as well as consultations with colleagues from ATLANTIS and wider research consortiums. In addition, the above stake- and rightholder engagement was employed, to identify key terms for the review that are relevant to actors involved in the Northern Spanish forestry sector.

The search terms were established to identify meta-analyses which have scrutinised the effects of forestry management practices on forest ecosystem and soil health. Keywords were selected to balance the output of the search to include all sources of relevance (sensitivity), while aiming to make the search as precise as possible to limit the amount of irrelevant output (specificity). For literature reviews in ecology, Pullin and Stewart (2006) recommend high-sensitivity, low-specificity searches, to assure repeatability of the search, because keywords and methodologies are not defined and used with the same rigour as common in medicine and health sciences.

For this purpose, an initial search string of relevant search terms was created, which is firstly tested in Web of Science for sensitivity and specificity. Obvious shortcomings (such as resulting in far too many or too few, or a large quantity of irrelevant references from different disciplines) are corrected by editing and refining the search string. When these shortcomings are resolved, the search string

will be tested by reviewing the title, abstracts and keywords of 100 randomly selected references (Foo et al., 2021) and to assess, whether the search keywords satisfactorily identified relevant sources. Foo et al. (2021) recommend to aim for a hit rate of 10% of the returned references to be relevant for the search (more on that below in the section on *Protocol comprehensiveness*). Finally, the comprehensiveness of the search keywords will be tested using the core set of previously identified benchmark studies (see below).

As we commonly observe that details on the comparators, outcomes and geographical scope/context are not specifically mentioned in the publications' title and abstracts, we chose to limit the search terms to the population, interventions and study type elements, as outlined below (Table 1). Search strings for SCOPUS, CAB Forest Science and Google Scholar are attached in Appendix III. Please note that due to the limitation of Google Scholar to search strings of max. 256 characters, the search string for this engine was accordingly shortened and simplified (Gusenbauer & Haddaway, 2020).

Table 1: Search string components for the PICO elements in the format for Web of Science. Given the information expected to be available in title, abstract and keywords, components were limited to the Population, Intervention and Study Type elements.

PICO element	Search string component
POPULATION	TS=(forest* OR (plantation NEAR/5 (forest* OR tree OR timber OR stand*)) OR (stand NEAR/5 (forest* OR tree OR timber)) OR (soil NEAR/5 (forest* OR tree OR timber OR stand*)))
INTERVENTION	TS=(silvicultur* OR forestry OR "managed forest*" OR harvest* OR (management NEAR/5 (forest* OR tree OR timber OR stand*)))
COMPARISON	NONE
OUTCOME	NONE
CONTEXT	NONE
STUDY TYPE	TS=(meta-analys* OR "meta analys*" OR "systematic review" OR "systematic literature review" OR (synthes* NEAR/5 (evidence OR research OR quantitative)))

Publications are only included into the review, if they were published in English. This is due to the lack of capacity to translate and include studies in other languages. However, we acknowledge that this may pose a language bias in the output of the study (Joanna Briggs Institute, 2014; Livoreil et al., 2017). Studies resulting from the review in other languages will therefore be recorded separately in a database and will be available for future analysis and referencing.

Protocol comprehensiveness

To test the selection of keywords and the scope of the search on sensitivity and specificity, we created a test list of benchmark studies that fulfilled the inclusion criteria to see how many of them are captured by the search (Foo et al., 2021; Table 2). These studies were collected before the review through external expert recommendations, relevant references of identified studies, and the research platform *ResearchRabbit* that allowed the random identification of studies without using search terms (Human Intelligence Technologies Incorporated, 2021).

Table 2: Test list of selected benchmark studies for the umbrella review. All studies fulfil the inclusion criteria. The benchmark studies serve to guide the umbrella review and to assess whether the search terms are well framed to identify key studies.

Reference	Forestry practice (intervention)	Forest ecosystem or soil health indicator (outcome)	Geographical scope (context)
Clarke et al. (2021)	Intensive/whole-tree harvest (WTH) or conventional/stem-only harvest (SO)	Soil organic carbon (SOC) and total N (TN) stocks; other soil nutrients (P, K, Ca, Mg, Zn, Na, Al, Mn); soil pH, exchangeable acidity (EA), cation exchange capacity (CEC) and base saturation (BS)	Temperate (Europe)
Rehschuh et al. (2021)	Monocultures (beech, conifers) or mixed stands	Soil organic carbon (SOC) and total N (TN) stocks	Temperate (Europe)
Dieler et al. (2017)	(1) managed (thinning, thinning from below, shelterwood, group selection logging, selection forest systems) or unmanaged stands; (2) mixed or pure stands	(1) forest species richness/diversity; (2) stand productivity	Temperate (Europe)
Hume et al. (2018)	Intensive (WTH) or conventional (SO) biomass harvesting	Soil C, N and P stocks	Temperate and boreal (North America, Europe)
Chaudhary et al. (2016)	Managed (clear-cutting, retention, selection and retention systems for temperate and boreal) or unmanaged stands	Forest species richness	Global (but effect sizes for specific management practices for temperate and boreal forests specifically)
Achat et al. (2015)	Intensive (WTH) or conventional (SO) biomass harvesting	Soil organic carbon (SOC) at different layers	Global (effect sizes for temperate and boreal forests separate)
Achat et al. (2015)	Intensive (WTH) or conventional (SO) biomass harvesting	Soil nutrient outputs (SOC; total and available N, P, K, Ca, Mg), wide range of chemical and biological soil fertility (CEC, BS, pH, EA, etc.)	Global (91% of sites in temperate and boreal climate)
Fedrowitz et al. (2014)	Retention systems or clear-cuts or unharvested forests	Forest species richness and abundance	Temperate and boreal (North America, Europe)
Mori & Kitagawa (2014)	Retention systems or primary forests	Forest species richness and abundance	Global (effect sizes for temperate and boreal forests separate)

Duguid & Ashton (2013)	Even-aged regeneration methods (clearcut, seed tree, shelterwood) or uneven-aged management (selection); thinning treatments	Understory plant species diversity	Temperate (North America)
Jerabkova et al. (2011)	Retention or clearcut harvest	Soil nitrogen fluxes	Temperate and boreal (North America, Europe, Japan)
Nave et al. (2010)	Hardwood or coniferous or mixed species composition; harvesting type (clearcut or thinning); if possible, harvesting intensity (WTH or stem-only)	Soil carbon at different layers	Temperate (global)
Paillet et al. (2010)	Management intensity (clearcut, selective felling)	Species richness (vascular plants, bryophytes, lichens, birds, saproxylic beetles, fungi)	Temperate (Europe)
Zwolak (2009)	Retention or clearcut harvest or undisturbed forest	Abundance of small mammals (deer mice, red-backed voles)	Temperate (North America)
Rosenvald & Lõhmus (2008)	Retention or clearcut harvest	Forest species richness and abundance	Temperate, boreal, subtropical (mostly North America, Europe)
Vanderwel et al. (2007)	Retention harvest at different intensity levels	Abundance of 14 bird species	Temperate (North America)

In addition to the benchmark studies, to test the scope of search keywords and to estimate the final number of relevant studies for time and resource management, the above outlined pilot screening will be conducted based on 100 randomly selected references (Foo et al., 2021). For this purpose, title and abstracts, and, later on, full-text manuscripts of these references will be screened using the below outlined review process. The percentage of included studies will then be used to estimate the total number of records that the review will result in. This percentage or "hit rate" is recommended by Foo et al. (2021) to be at around 10% of the search output.

Review process

Screening stages

The umbrella review will be structured according to the four stages of the PRISMA guidelines (Moher et al., 2009): identification, screening, eligibility and inclusion.

1. IDENTIFICATION: In the first stage of the review, all articles eligible for screening are identified, by running the above defined search strings in the search engines. Manuscript duplicates between databases are possible.

2. SCREENING: The initial screening will use the output from the first stage and will be limited to the review of titles, abstracts and keywords of the manuscripts (*title and abstract screening*). Concluding this phase, all articles are retained for full-text review when they either fulfil the above outlined inclusion criteria or when it is unclear whether they fulfil the criteria.
3. ELIGIBILITY: In the second review stage, the manuscripts from the screening stage are reviewed at full-text stage (*full-text review*). Articles are selected, if they fulfil all inclusion criteria, if data is available and not duplicated from earlier publications (see chapter *In- and exclusion criteria*). Reasons for article exclusion at this stage will be documented using the ROSES checklist and flow diagram and reported in the appendix of the publication (Collaboration for Environmental Evidence, 2018; N.R. Haddaway et al., 2018). Finally, the quality of the remaining reviews will be critically appraised and studies with very low variability scores excluded (see details below).
4. INCLUSION: In this final stage, articles filtered and selected in the eligibility stage are recorded in a database and data is extracted.

At the end of the identification stage, all resulting references will be downloaded as .ris files. For Google Scholar, the R package *GSscraper* will be used to download the first 300 references (Haddaway, 2022). Afterwards, the results will be uploaded and merged in the online review tool *CADIMA* to remove duplicates between databases and to conduct the following article synthesis (Kohl et al., 2018b, 2018a). This enables us to ensure a coherent, detailed and transparent review process. In addition, results at individual review stages will be visualised using the ROSES flow diagram (see Appendix II; Haddaway et al., 2018). The decision trees for the two screening procedures are attached in Appendix IV.

Finally, to avoid conflicts of interests, review team members who are co-authors of articles found during the search or who are close collaborators with an author, will not review the respective publication for inclusion and delegate the decision to unbiased team members.

Search consistency check

As a first step to test protocol comprehensibility, search terms and inclusion criteria were revised and discussed by the review team during the design phase, to be understood as unanimously as possible. In the review process, to ensure reproducibility of the review and consistency of results, nominal agreement between independent reviewers about the inclusion of studies will then be quantified using Cohen's Kappa statistic (Collaboration for Environmental Evidence, 2018). For this purpose, the title and abstract screening of the review process will be conducted by two members of the research team for a 10% random sample of the merged database search results (Woodcock et al., 2017). The level of agreement between the reviewers *A* and *B* is calculated as a Kappa score *K*, as a function of the observed agreement P_o and expected agreement P_e (Cohen, 1960):

$$K = \frac{\mathrm{P_0 - P_e}}{\mathrm{1 - P_e}} \quad (1)$$

With P_o being the sum of observed agreement propensity between *A* and *B* on the share of publications to be included ($Incl_{AB}$) and excluded ($Excl_{AB}$) from all publications *N* (Warrens, 2015):

$$P_0 = \frac{(Incl_{AB} + Excl_{AB})}{N} \quad (2)$$

And P_e being the expected agreement between the reviewers for each of the categories, using the observed data to estimate the chance of both reviewers independently classifying studies as included (A_{incl} and B_{incl} respectively) or excluded (A_{excl} and B_{excl} respectively):

$$P_e = \frac{(A_{incl})}{N}\frac{(B_{incl})}{N} + \frac{(A_{excl})}{N}\frac{(B_{excl})}{N} \quad (3)$$

Kappa scores around 0 indicate that any agreements between reviewers are considered to be by chance. The highest possible score of 1 in turn shows that reviewers are in full agreement on the choices made (Cohen, 1960). Kappa scores above 0.6 are considered substantial, scores above 0.8 as almost perfect (Landis & Koch, 1977). Thus, if *K* is lower than 0.6 for our sample, we will discuss and review disagreements between choices between reviewers and to repeat the process of sampling 10% of the search results. Kappa scores will then be calculated again and this process will be repeated until agreement between reviewers is at least substantial (*K > 0.6*).

Critical appraisal

The number of quantitative evidence reviews in environmental sciences and management has increased significantly in order to synthesise the vast number of primary studies published and inform decision-making (Pullin et al., 2022). However, not all syntheses apply methodological standards for evidence review with the same rigour and biases in reviewing and reporting results are common (O'Leary et al., 2017). This can have important consequences for policy conclusions drawn from these syntheses, which has motivated synthesis specialists, such as the Collaboration for Environmental Evidence (CEE) to develop standardised guidelines for assessing the reliability and methodological quality of evidence reviews and meta-analyses (Woodcock et al., 2014). These quality assessments are now commonly required in evidence synthesis (e.g. Foo et al., 2021; Haddaway et al., 2018; Woodcock et al., 2017). Importantly, these risks need to be considered for umbrella reviews, as biases from primary studies can be transferred to this meta-level if they are not recognised in meta-analyses due to a lack of critical appraisal. Therefore, critical appraisals of meta-analyses included in these reviews are already a common requirement in guidelines for umbrella reviews in health and medical sciences (Aromataris et al., 2015; Fusar-Poli & Radua, 2018; Joanna Briggs Institute, 2014).

For this purpose, we will adopt the standardised Collaboration for Environmental Evidence Synthesis Appraisal Tool (CEESAT) for evidence reviews to score the methodological quality of the meta-analyses that are identified in our umbrella review (Collaboration for Environmental Evidence, 2020; Woodcock et al., 2014). The CEESAT was developed to assess the rigour and transparency of the methodology of a review. In total, each meta-analysis is evaluated against a checklist of 16 criteria for the different stages of the review, such as methods, review procedures, critical appraisal, data extraction and limitations (for the checklist questions, see Appendix V). Each criterion is rated on a four-point ordinal scale, from red (poor quality), via amber and green to gold (high quality). The authors of CEESAT advise against the translation of the colour codes into numerical scores for the assessment of overall review reliability to avoid oversimplification (Collaboration for Environmental Evidence, 2021). Therefore, to be able to estimate the quality performance of the identified studies and exclude those that are at risk of having very poor reliability, we created five broad categories of review validity (Table 3).

Table 3: Study validity categories based on the colour coding from CEESAT for critical appraisal of study quality. Each study receives a total of 16 scores that can be RED, AMBER, GREEN or GOLD. Categorisation should start with the 'very low' category and move down the list to higher categories, if conditions do not apply.

VALIDITY CATEGORY	The review scores…
Very low validity:	RED in any of the key criteria 1.1, 3.1, 3.2, 4.1, 6.1, 7.1 or 7.2. *OR* RED in at least 8 criteria *AND* AMBER in all remaining criteria.
Low validity:	GOLD and/or GREEN in up to 4 criteria.*
Medium validity:	GOLD and/or GREEN in 5 to 8 criteria.*
High validity:	GOLD and/or GREEN in 9 to 12 criteria.*
Very high validity:	GOLD and/or GREEN in at least 13 criteria.*
	* The remaining criteria are scored AMBER and/or RED, but cannot be RED in any of the key criteria outlined in '*Very low*'.

When all 16 criteria have been scored, each colour is counted and the validity category identified by starting with the 'very low' category and moving down the list to identify the category where the study meets the conditions. The accuracy and repeatability of the categorisation process was tested and adjusted prior to the review: firstly, on the benchmark studies specified in Table 2; and secondly, on a dataset of another umbrella review by the authors (PAM, CM, JCY; in progress). To improve the synthesis of clear, reliable findings in the full review, meta-analyses with poor methodological quality (*very low validity*) will be excluded from the subsequent data synthesis. This also implies that reviews with red scores for any of the criteria that we consider of high relevance for review reliability (1.1, 3.1, 3.2, 4.1, 6.1, 7.1, 7.2) will be excluded. We furthermore decided to add a prerequisite to criterion 3.2, which states that reviews which rely on Google Scholar as their sole search engine are equally scored as RED, as Google Scholar is not recommended as a stand-alone search engine due to a lack of systematic outputs (Foo et al., 2021; Gusenbauer & Haddaway, 2020). Reasons for overall judgement of reviews will be transparently documented in the critical appraisal Excel sheet attached in the Supplementary Materials of the review. For grey literature, we will furthermore adopt the AACODS checklist (Authority, Accuracy, Coverage, Objectivity, Date, Significance), developed for evidence reviews in health science as a tool to critically appraise the origin of these studies (Tyndall, 2010). For the AACODS checklist, please see Appendix VI.

Finally, as recommended in the JBI and CEE guidelines, the critical appraisal process will be carried out independently by two reviewers for a sub-sample of at least 10% of the included records, to verify procedural consistency (Collaboration for Environmental Evidence, 2018; Joanna Briggs Institute, 2014). Kappa scores will be calculated to assess the level of agreement between reviewers. The critical appraisal process will be essential in answering the third research question on the quality and reliability of evidence from existing meta-analyses and syntheses on impacts of forestry practices on forest and soil health.

Data extraction

The meta-analyses that are selected after the full-text screening and critical appraisal will be used to address the research questions on the effect of forest management practices on forest ecosystem and soil health. We will distinguish between primary data, which will be extracted in detail, and secondary data, which is mainly used to provide a context and overview for the subsequent data synthesis (adapted from Aromataris et al., 2015; Joanna Briggs Institute, 2014).

Primary data that will be extracted in detail includes:

- *Study citation details*: authors and title, publication year and journal
- PICO elements:
    - *Population* details: forest ecosystem characteristics, sample size n
    - *Interventions*: type, frequency and intensity
    - *Comparators*: number and type of compared ecosystems
    - *Outcomes*: types, direction and intensity
- *Environmental context* or geographical scope of the study
- *Type of study*: type of quantitative evidence synthesis, such as random or fixed effects meta-analysis, meta-aggregative synthesis, etc.
- Key findings:
    - *Effect sizes*: the number and direction of summary effect sizes reported for each synthesis
    - Measures of *heterogeneity* (if reported): it will probably be the I^2 statistic, indicating the variance between effect sizes of included studies (expressed in percentage, with values being >75% high, 51-75% medium and ≤50% low variance; see Senior et al., 2016)[1]

Additional, **secondary data** that will be extracted includes:

- *Objectives* and general aims of the synthesis
- *Inclusion criteria* of the synthesis, if reported
- *Databases* sourced and searched
- *Search details* of synthesis: search terms or search strings, if reported
- Details of primary studies included in the synthesis
    - *Number* of studies included
    - *Country* of origin
    - *Year range* of studies
    - *Type* of primary studies (e.g. laboratory or greenhouse experiment, field study), if reported
    - *Study design* of primary studies (e.g. before-and-after; randomised controlled studies), if reported
- Other relevant *results*, including the *significance* and *direction* of effects per outcome
- *Critical appraisal tool* used in the synthesis for the primary studies, with details on
    - *Types of biases* assessed
    - *Results/quality ratings*
- *Method of synthesis* or analysis employed to synthesize the evidence: meta-analytical model
- *Location of all data* within the publication

[1] Where available, the I^2 statistic is expected to indicate rather high degrees of heterogeneity, as reported for instance in Senior et al. (2016) for studies in ecology and evolution.

- Any *comments* or notes regarding any included study

The data will be extracted from tables, figures, abstracts and the main text. Graphical data from figures will be extracted using the R package *metaDigitise* (Pick et al., 2018). If any of the data cannot be retrieved from the publication, authors of meta-analyses will be contacted. Data from primary studies will not be retrieved or reported, unless a particular outcome is informed by a single included study (Aromataris et al., 2015). Data extraction forms will be provided in the appendix of the review.

As in previous elements of the review, data extraction will be conducted by two reviewers independently for a subset of 10% of the data to ensure the consistency of the data extraction procedure (Collaboration for Environmental Evidence, 2018).

Data synthesis

The main aim of an umbrella review is to summarize the evidence synthesised by a large number of reviews and to assess the quality of the available data. As outlined in the introduction, the goal is therefore not to further synthesise the data obtained, but to provide an overview of the existing available data and make the evidence more readily and transparently available to practitioners and policy makers (Aromataris et al., 2015; Joanna Briggs Institute, 2014).

In this review, we will provide narrative, tabular, and graphic summaries of the results with the aim to answer the research questions (RQ) posed prior to the review. Firstly, an overview of the article screening process and the available literature will be given, which will be accompanied by the ROSES flow diagram for a detailed overview of the in- and exclusion of studies at each step of the process (Haddaway et al., 2018). Reasons for exclusion of articles in the full-text review stage will be detailed in the supplementary material of the publication. The geographical distribution of primary study data will be visualised using the R packages *ggplot2* and *ggbiome* (Stefan & Levin, 2022; Wickham, 2016).

Secondly, overall results and descriptive details will be communicated to assess the internal and external validity/relevance of results, to highlight important outcomes and points of interest (Collaboration for Environmental Evidence, 2018). The narrative description of results will be supported by tables, giving a summary of the included studies, their citation details, PICO elements, effect modifiers that influence results, etc. Graphs such as forest plots and heatmaps will be used to visualise effect sizes[2] and heterogeneity between and within studies (RQ 1). To support the graphical representation of overall results and identify evidence synthesis gaps (RQ 2), we will employ evidence review mapping (ERM) proposed by O'Leary et al. (2017). ERM is a complementary visualisation approach to systematic reviews and can be used to highlight actual and apparent (cryptic) evidence gaps, as well as redundancy (various reviews on a similar topic or question) in systematic reviews and meta-analyses (O'Leary et al., 2017; Woodcock et al., 2017). The critical appraisal of study quality will equally be summarised in the form of tables and heatmaps, to highlight which criteria are given more or less attention (RQ 3).

Parallel to the production of a scientific publication based on this data synthesis, we will use the ERM and other graphics for the production of a short summary for practitioners. It will be used to

[2] The direction of intervention effects is indicated by a 'traffic light' indicator, with green, amber and red colours specifying positive/beneficial, neutral or negative/detrimental effects of a management practice on a forest health indicator respectively (Joanna Briggs Institute, 2014).

circulate the findings of the study to existing and future collaborating stake- and rightholders of the project and to provide recommendations for decision-making.

Discussion and conclusion

As a final step of the umbrella review, data synthesised and summarised from selected quantitative evidence reviews will be discussed and set into context of existing literature, policy and practice recommendations (Aromataris et al., 2015). Furthermore, they will be used to answer the initially posed research questions, to identify knowledge gaps and discuss steps forward for future evidence reviews. This includes a thorough discussion of the quality and reliability of data resulting from these meta-analyses and the study limitations that are associated with biases, which were identified in the process (O'Leary et al., 2017). These can be related to systematic errors conducted during study selection and data extraction, which can hamper transparency or repeatability of processes, or (the lack of) critical appraisal procedures (Joanna Briggs Institute, 2014).

Finally, umbrella reviews are only as strong as its input from included evidence reviews (Pullin et al., 2022). Therefore, the discussion section will also scrutinise and critically reflect limitations and possible biases in the umbrella review process, and outline how these limitations were intended to be overcome. This can for instance be done by employing strict critical appraisal methods and transparently documenting decisions on in- and exclusion of meta-analyses.

To complete the review, conclusions will be drawn and answers to the posed research questions reiterated. If appropriate, recommendations for future reviews, but also policies and decision-making will be posed here.

Competing interests

No competing interests have been identified. The umbrella review is part of the coordinated project *ATLANTIS* (Development of soil-smart forestry practices and of early vulnerability diagnosis tools to improve soil conservation and long-term stability of Iberian Atlantic Forests), funded by the Spanish Ministry of Science and Innovation.

References

Achat, D. L., Deleuze, C., Landmann, G., Pousse, N., Ranger, J., & Augusto, L. (2015). Quantifying consequences of removing harvesting residues on forest soils and tree growth - A meta-analysis. *Forest Ecology and Management*, *348*, 124–141. https://doi.org/10.1016/j.foreco.2015.03.042

Achat, David L., Fortin, M., Landmann, G., Ringeval, B., & Augusto, L. (2015). Forest soil carbon is threatened by intensive biomass harvesting. *Scientific Reports*, *5*, 1–10. https://doi.org/10.1038/srep15991

Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-Based Healthcare*, *13*(3), 132–140. https://doi.org/10.1097/XEB.0000000000000055

Castaño-Villa, G. J., Estevez, J. V., Guevara, G., Bohada-Murillo, M., & Fontúrbel, F. E. (2019). Differential effects of forestry plantations on bird diversity: A global assessment. *Forest Ecology and Management*, *440*(December 2018), 202–207. https://doi.org/10.1016/j.foreco.2019.03.025

Chaudhary, A., Burivalova, Z., Koh, L. P., & Hellweg, S. (2016). Impact of Forest Management on Species Richness: Global Meta-Analysis and Economic Trade-Offs. *Scientific Reports*, *6*, 1–11. https://doi.org/10.1038/srep23954

Clarke, N., Kiær, L. P., Janne Kjønaas, O., Bárcena, T. G., Vesterdal, L., Stupak, I., Finér, L., Jacobson, S., Armolaitis, K., Lazdina, D., Stefánsdóttir, H. M., & Sigurdsson, B. D. (2021). Effects of intensive biomass harvesting on forest soils in the Nordic countries and the UK: A meta-analysis. *Forest Ecology and Management*, *482*. https://doi.org/10.1016/j.foreco.2020.118877

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*, 37–46. https://doi.org/10.1177/001316446002000104

Collaboration for Environmental Evidence. (2018). *Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.0 (AS Pullin, GK Frampton, B Livoreil & G Petrokofsky, Eds)*. Retrieved on 15 August 2022, from https://environmentalevidence.org/information-for-authors/

Collaboration for Environmental Evidence. (2020). *The Collaboration for Environmental Evidence Synthesis Appraisal Tool (CEESAT) for Overviews. Version 2.1*. https://environmentalevidence.org/ceeder/about-ceesat/

Collaboration for Environmental Evidence. (2021). *CEE Database of Evidence Reviews (CEEDER)*. Services for Evidence Users. Retrieved on 14 August 2022, from https://environmentalevidence.org/ceeder/

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*(4), 739–766. https://doi.org/10.1007/s10342-017-1056-1

Duguid, M. C., & Ashton, M. S. (2013). A meta-analysis of the effect of forest management for timber on understory plant species diversity in temperate forests. *Forest Ecology and Management*, *303*, 81–90. https://doi.org/10.1016/j.foreco.2013.04.009

Fedrowitz, K., Koricheva, J., Baker, S. C., Lindenmayer, D. B., Palik, B., Rosenvald, R., Beese, W., Franklin, J. F., Kouki, J., Macdonald, E., Messier, C., Sverdrup-Thygeson, A., & Gustafsson, L.

(2014). Can retention forestry help conserve biodiversity? A meta-analysis. *Journal of Applied Ecology*, *51*(6), 1669–1679. https://doi.org/10.1111/1365-2664.12289

Foo, Y. Z., O'Dea, R. E., Koricheva, J., Nakagawa, S., & Lagisz, M. (2021). A practical guide to question formation, systematic searching and study screening for literature reviews in ecology and evolution. *Methods in Ecology and Evolution*, *12*(9), 1705–1720.

Fusar-Poli, P., & Radua, J. (2018). Ten simple rules for conducting umbrella reviews. *Evidence-Based Mental Health*, *21*(3), 95–100. https://doi.org/10.1136/ebmental-2018-300014

Gazol, A., Hereş, A. M., & Curiel Yuste, J. (2021). Land-use practices (coppices and dehesas) and management intensity modulate responses of Holm oak growth to drought. *Agricultural and Forest Meteorology*, *297*(September). https://doi.org/10.1016/j.agrformet.2020.108235

Godin, K., Stapleton, J., Kirkpatrick, S. I., Hanning, R. M., & Leatherdale, S. T. (2015). Applying systematic review search methods to the grey literature: A case study examining guidelines for school-based breakfast programs in Canada. *Systematic Reviews*, *4*(1), 1–10. https://doi.org/10.1186/s13643-015-0125-0

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181–217. https://doi.org/10.1002/jrsm.1378

Haddaway, N. (2022). *GSscraper: An R package and Shiny app for exporting search results from Google Scholar. Zenodo.* 10.5281/zenodo.6458134

Haddaway, N. R., Kohl, C., Rebelo Da Silva, N., Schiemann, J., Spök, A., Stewart, R., Sweet, J. B., & Wilhelm, R. (2017). A framework for stakeholder engagement during systematic reviews and maps in environmental management. *Environmental Evidence*, *6*(1). https://doi.org/10.1186/s13750-017-0089-8

Haddaway, N.R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of google scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, *10*(9), 1–17. https://doi.org/10.1371/journal.pone.0138237

Haddaway, N.R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES Reporting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*(1), 4–11. https://doi.org/10.1186/s13750-018-0121-7

Haddaway, Neal R., & Pullin, A. S. (2014). The Policy Role of Systematic Reviews: Past, Present and Future. *Springer Science Reviews*, *2*(1–2), 179–183. https://doi.org/10.1007/s40362-014-0023-1

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*(6160), 850–853. https://doi.org/10.1126/science.1244693

Hereş, A. M., Petritan, I. C., Bigler, C., Curtu, A. L., Petrea, Ş., Petritan, A. M., Polanco-Martínez, J. M., Rigling, A., & Curiel Yuste, J. (2021). Legacies of past forest management determine current responses to severe drought events of conifer species in the Romanian Carpathians. *Science of the Total Environment*, *751*, 1–11. https://doi.org/10.1016/j.scitotenv.2020.141851

Hill, S. L. L., Arnell, A., Maney, C., Butchart, S. H. M., Hilton-Taylor, C., Ciciarelli, C., Davis, C., Dinerstein, E., Purvis, A., & Burgess, N. D. (2019). Measuring Forest Biodiversity Status and

Changes Globally. *Frontiers in Forests and Global Change*, *2*(November), 1–11. https://doi.org/10.3389/ffgc.2019.00070

Hua, F., Adrian Bruijnzeel, L., Meli, P., Martin, P. A., Zhang, J., Nakagawa, S., Miao, X., Wang, W., McEvoy, C., Peña-Arancibia, J. L., Brancalion, P. H. S., Smith, P., Edwards, D. P., & Balmford, A. (2022). The biodiversity and ecosystem service contributions and trade-offs of forest restoration approaches. *Science*, *376*(6595), 839–844. https://doi.org/10.1126/science.abl4649

Human Intelligence Technologies Incorporated. (2021). *ResearchRabbit*. Retrieved on 19 May 2022, from https://researchrabbitapp.com/

Hume, A. M., Chen, H. Y. H., & Taylor, A. R. (2018). Intensive forest harvesting increases susceptibility of northern forest soils to carbon, nitrogen and phosphorus loss. *Journal of Applied Ecology*, *55*(1), 246–255. https://doi.org/10.1111/1365-2664.12942

Ibáñez, I., Acharya, K., Juno, E., Karounos, C., Lee, B. R., McCollum, C., Schaffer-Morrison, S., & Tourville, J. (2019). Forest resilience under global environmental change: Do we have the information we need? A systematic review. *PLoS ONE*, *14*(9), 1–17. https://doi.org/10.1371/journal.pone.0222207

Jerabkova, L., Prescott, C. E., Titus, B. D., Hope, G. D., & Walters, M. B. (2011). A meta-analysis of the effects of Clearcut and variable-retention harvesting on soil nitrogen fluxes in boreal and temperate forests. *Canadian Journal of Forest Research*, *41*(9), 1852–1870. https://doi.org/10.1139/x11-087

Joanna Briggs Institute. (2014). The Joanna Briggs Institute Reviewers' Manual 2014: Methodology for JBI Umbrella Reviews. *Joanna Briggs Institute*, 1–34. http://joannabriggs.org/assets/docs/sumari/ReviewersManual_Mixed-Methods-Review-Methods-2014-ch1.pdf

Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018a). Correction to: Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools (Environmental Evidence (2018) 7 (8) DOI: 10.1186/s13750-018-0115-5). *Environmental Evidence*, *7*(1), 13750. https://doi.org/10.1186/s13750-018-0124-4

Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018b). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, *7*(1), 1–17. https://doi.org/10.1186/s13750-018-0115-5

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data Data for Categorical of Observer Agreement The Measurement. *Biometrics*, *33*(1), 159–174.

Lázaro-Lobo, A., Ruiz-Benito, P., Lara-Romero, C., & Castro-Díez, P. (2022). Biotic, abiotic, and anthropogenic drivers of demographic performance of non-native Eucalyptus and Pinus species in forested areas of Spain. *Forest Ecology and Management*, *510*. https://doi.org/10.1016/j.foreco.2022.120111

Leuschner, C., Förster, A., Diers, M., & Culmsee, H. (2022). Are northern German Scots pine plantations climate smart? The impact of large-scale conifer planting on climate, soil and the water cycle. *Forest Ecology and Management*, *507*, 120013. https://doi.org/10.1016/j.foreco.2022.120013

Livoreil, B., Glanville, J., Haddaway, N. R., Bayliss, H., Bethel, A., De Lachapelle, F. F., Robalino, S., Savilaakso, S., Zhou, W., Petrokofsky, G., & Frampton, G. (2017). Systematic searching for

environmental evidence using multiple tools and sources. *Environmental Evidence*, *6*(1). https://doi.org/10.1186/s13750-017-0099-6

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Mori, A. S., & Kitagawa, R. (2014). Retention forestry as a major paradigm for safeguarding forest biodiversity in productive landscapes: A global meta-analysis. *Biological Conservation*, *175*, 65–73. https://doi.org/10.1016/j.biocon.2014.04.016

Naudts, K., Chen, Y., McGrath, M. J., Ryder, J., Valade, A., Otto, J., & Luyssaert, S. (2016). Forest management: Europe's forest management did not mitigate climate warming. *Science*, *351*(6273), 597–599. https://doi.org/10.1126/science.aad7270

Nave, L. E., Vance, E. D., Swanston, C. W., & Curtis, P. S. (2010). Harvest impacts on soil carbon storage in temperate forests. *Forest Ecology and Management*, *259*(5), 857–866. https://doi.org/10.1016/j.foreco.2009.12.009

O'Leary, B. C., Woodcock, P., Kaiser, M. J., & Pullin, A. S. (2017). Evidence maps and evidence gaps: Evidence review mapping as a method for collating and appraising evidence reviews to inform research and policy. *Environmental Evidence*, *6*(19). https://doi.org/10.1186/s13750-017-0096-9

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*(1), 879–910. https://doi.org/10.17705/1cais.03743

Paillet, Y., Bergès, L., Hjältén, J., Ódor, P., Avon, C., Bernhardt-Römermann, M., Bijlsma, R. J., De Bruyn, L., Fuhr, M., Grandin, U., Kanka, R., Lundin, L., Luque, S., Magura, T., Matesanz, S., Mészáros, I., SebastiÀ, M. T., Schmidt, W., Standovár, T., … Virtanen, R. (2010). Biodiversity differences between managed and unmanaged forests: Meta-analysis of species richness in Europe. *Conservation Biology*, *24*(1), 101–112. https://doi.org/10.1111/j.1523-1739.2009.01399.x

Pan, Y., Birdsey, R. A., Phillips, O. L., & Jackson, R. B. (2013). The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 593–622. https://doi.org/10.1146/annurev-ecolsys-110512-135914

Perring, M. P., Bernhardt-Römermann, M., Baeten, L., Midolo, G., Blondeel, H., Depauw, L., Landuyt, D., Maes, S. L., De Lombaerde, E., Carón, M. M., Vellend, M., Brunet, J., Chudomelová, M., Decocq, G., Diekmann, M., Dirnböck, T., Dörfler, I., Durak, T., De Frenne, P., … Verheyen, K. (2018). Global environmental change effects on plant community composition trajectories depend upon management legacies. *Global Change Biology*, *24*(4), 1722–1740. https://doi.org/10.1111/gcb.14030

Pick, J. L., Nakagawa, S., & Noble, D. W. A. (2018). *Reproducible, flexible and high-throughput data extraction from primary literature: The metaDigitise R package*. Biorxiv. https://doi.org/10.1101/247775

Pullin, A. S., Cheng, S. H., Jackson, J. D., Eales, J., Envall, I., Fada, S. J., Frampton, G. K., Harper, M., Kadykalo, A. N., Kohl, C., Konno, K., Livoreil, B., Ouédraogo, D.-Y., O'Leary, B. C., Pullin, G., Randall, N., Rees, R., Smith, A., Sordello, R., … Woodcock, P. (2022). Standards of conduct and reporting in evidence syntheses that could inform environmental policy and management decisions. *Environmental Evidence*, *11*(1), 1–11. https://doi.org/10.1186/s13750-022-00269-9

Pullin, A. S., & Stewart, G. B. (2006). Guidelines for systematic review in conservation and environmental management. *Conservation Biology*, *20*(6), 1647–1656. https://doi.org/10.1111/j.1523-1739.2006.00485.x

Rehschuh, S., Jonard, M., Wiesmeier, M., Rennenberg, H., & Dannenmann, M. (2021). Impact of European Beech Forest Diversification on Soil Organic Carbon and Total Nitrogen Stocks–A Meta-Analysis. *Frontiers in Forests and Global Change*, *4*(February). https://doi.org/10.3389/ffgc.2021.606669

Rosenvald, R., & Lõhmus, A. (2008). For what, when, and where is green-tree retention better than clear-cutting? A review of the biodiversity aspects. *Forest Ecology and Management*, *255*(1), 1–15. https://doi.org/10.1016/j.foreco.2007.09.016

Secretariat of the Convention on Biological Diversity. (2010). *Global Biodiversity Outlook 3*. SCBD, Montréal.

Senior, A. M., Grueber, C. E., Kamiya, T., Lagisz, M., O'Dwyer, K., Santos, E. S. A., & Nakagawa, S. (2016). Heterogeneity in Ecological and Evolutionary Meta-Analyses: its Magnitudes and Implications. *Ecology*, *97*(12), 3293–3299. https://doi.org/10.1002/ecy.1591

Stefan, V., & Levin, S. (2022). *plotbiomes: Plot Whittaker biomes with ggplot2* (R package version 0.0.0.9001.).

Thompson, I., Mackey, B., McNulty, S., & Mosseler, A. (2009). Forest Resilience, Biodiversity, and Climate Change. A synthesis of the biodiversity/resilience/stability relationship in forest ecosystems. In *CBD Technical Series: Vol. no. 43* (Issue January). Secretariat of the Convention on Biological Diversity, Montreal. http://www.cbd.int/doc/publications/cbd-ts-43-en.pdf

Tyndall, J. (2010). *AACODS Checklist.* Flinders University. Retrieved on 24 May 2022, from http://dspace.flinders.edu.au/dspace/

Vanderwel, M. C., Malcolm, J. R., & Mills, S. C. (2007). A meta-analysis of bird responses to uniform partial harvesting across North America. *Conservation Biology*, *21*(5), 1230–1240. https://doi.org/10.1111/j.1523-1739.2007.00756.x

Wang, C., Zhang, W., Li, X., & Wu, J. (2022). A global meta-analysis of the impacts of tree plantations on biodiversity. *Global Ecology and Biogeography*, *31*(3), 576–587. https://doi.org/10.1111/geb.13440

Warrens, M. J. (2015). Five Ways to Look at Cohen's Kappa. *Journal of Psychology & Psychotherapy*, *5*(4), 8–11. https://doi.org/10.4172/2161-0487.1000197

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. https://ggplot2.tidyverse.org.

Woodcock, P., O'Leary, B. C., Kaiser, M., & Pullin, A. (2017). Your evidence or mine? Systematic evaluation of reviews of marine protected area effectiveness. *Fish and Fisheries*, *18*(4), 668–681.

Woodcock, P., Pullin, A. S., & Kaiser, M. J. (2014). Evaluating and improving the reliability of evidence syntheses in conservation and environmental science: A methodology. *Biological Conservation*, *176*, 54–62.

Zwolak, R. (2009). A meta-analysis of the effects of wildfire, clearcutting, and partial harvest on the abundance of North American small mammals. *Forest Ecology and Management*, *258*(5), 539–545. https://doi.org/10.1016/j.foreco.2009.05.033

Appendix I

Detailed lists of the indicators for INTERVENTION (Table A1) and OUTCOME (Table A2) elements used during the review.

Table A1: Categories of forest management practices (INTERVENTIONS) that are included in the meta-analysis. Sub-categories are indicated in italic and with indentation below the category they belong to.

Forest management practices	Description	Synonyms
Harvesting	Tree removal practices, commonly at the end of the rotation period	Logging
Clearcutting	Most or all trees are uniformly removed; traditionally only stems are removed and stumps retained	Clearfelling; stem-only harvesting (SH)
Seed-tree system	Removal of all except a few trees to supply seeds for the next crop; managed as a clearcut system	
Shelterwood system	Removal of all except a few trees to protect and shelter the developing regeneration	Tree shelter
Reduced impact logging (RIL)	Timber harvesting methods applied to reduce the negative environmental impact on forest plantations and soils; practiced in the tropical regions	
Intense biomass production and removal	Litter and residue removal for biomass and biofuel production; this can include the removal of stumps and whole trees	Short-rotation woody crops; coppice forestry; intense biomass harvesting; whole tree harvesting (WTH); stump harvest
Thinning	Selective removal of trees, commonly during the rotation period to improve growth or health	Tree removal
Single tree selection	Removal of individual trees to promote uneven-aged stands	
Group selection	Removal of small groups of trees to promote uneven-aged stands	
Retention system	Retaining of a certain percentage of individual trees or groups of trees for structural diversity; usually more trees than in shelterwood systems, creating an uneven-aged forest structure	
Tree species	Tree species present in the plantation	Plantation species; forest species
Native species	Plantations consisting of tree species that are indigenous to the region	Autochthon, indigenous species
Exotic species	Plantations consisting of tree species that were introduced to the region	Allochthon, non-native, alien, introduced species

Rotation length	Age of stand or trees since re- or afforestation	Stand age, tree age, forest age
Forest structure	Continuity of forest cover, density and structure of canopy layers and understory	Forest cover; mosaic forest structure
Even-aged forest	Forest stands of the same age class, with uniform height and diameter distributions	Uniform cover
Uneven-aged forest	Forest stands of different age classes, with differing height and diameter distributions	Continuous cover; multi-cohort stands; all-age stands
Forest type	Tree species composition and diversity in the forest stand	Stand type; monoculture; polyculture
Broadleaved forest	Forest stands dominated by angiosperm species, that are fruit-bearing hardwoods with flat leaves	Deciduous forest; hardwood forest; sometimes ever-green (but fewer species)
Coniferous forest	Forest stands dominated by gymnosperm species, that are cone-bearing softwoods, with needle- or scale-like leaves and usually ever-green	Needleleaved forest; Ever-green forest
Mixed forest	Forest stands with two or more dominant tree species; commonly a mix of both coniferous and broadleaved species	

Table A2: Categories of forest health indicators (OUTCOMES) that are included in the meta-analysis. Sub-categories are indicated in italic and with indentation below the category they belong to.

Forest health indicators	Description	Synonyms
FOREST ECOSYSTEM HEALTH		
Tree growth	Annual increase of the basal area of tree; sometimes also measured as increases in canopy cover (e.g. through the leaf area index/LAI) and ingrowth of trees (shrub-level trees that cross the threshold of basal area or height to adult trees)	Wood and plant biomass (incl. foliage); ingrowth
Productivity	Measurement of annual tree growth through (extractable) timber volume	Yield
Mortality	Measurement of annual tree volume lost by dead trees	Tree death
Defoliation	Relative deficit of leaves or needles on a tree crown (compared to a reference), which can indicate tree condition and problems such as illness, low productivity or plant stress	Foliage loss
Predators and pests	Predatory or parasitic animals and plants, affecting forest resistance and recovering capacity	Herbivory; fungal, bacterial, diseases; rusts; cankers; boring, chewing, sucking, foliage feeding insects; invasion; invasive species
Regeneration capacity	Capacity of the forest to regenerate naturally, measured through the survival of tree saplings, availability of seed banks, resprouters, etc.	Possibly ecological resilience; resprouting; recruitment
Ecological resilience	The overall capacity of an ecosystem to respond to and recover from disturbances	
Resilience	The capacity of an ecosystem to return to the previous ecological state after a disturbance or perturbation, measured as the performance *before* and *after* the disturbance took place	
Resistance	The capacity of an ecosystem to buffer or absorb disturbances or perturbations without being significantly altered, measured as the difference in performance *during* and *before* a disturbance	Persistence
Recovery	Process where an ecosystem returns to its previous ecological state, measured as the performance *during* and *after* a disturbance	

Relative resilience	The difference in growth between phases of resilience (*before-after*) and resistance (*before-during*) to a disturbance	Weighted resilience
Forest water cycle	Water processes above- and belowground	Soil moisture, water runoff, water infiltration, groundwater recharge, water use efficiency, water retention in leafs and litter, evapotranspiration
Ecosystem carbon cycle	Carbon contents aboveground, in vegetation, litter, and deadwood	
FOREST BIODIVERSITY		
Species diversity	Animal, plant (incl. tree and understory plant diversity), fungal and microbial diversity above- and belowground	Species richness
Species abundance	Number of individuals of animal or plant species above- and belowground	
SOIL HEALTH		
Soil nutrients	Key nutrients in soils that are essential for the growth and reproduction of plants	Plant nutrients
Soil carbon (C)	Contents of belowground carbon	Soil total carbon, soil organic carbon (SOC), soil microbial biomass carbon (MBC), soil carbon stock, carbon storage, carbon sequestration, soil respiration, soil CO2 efflux
Soil nitrogen (N)	Contents of belowground nitrogen	
Soil phosphorous (P)	Contents of belowground phosphorous	
Other soil nutrients	Contents of other important soil nutrients for plant growth and forest health	Soil methane (CH4), soil potassium (K), soil cadmium (Ca), soil greenhouse gases (GHG), soil microbial biomass nitrogen (MBN), foliar nutrients (N, P, Ca)
Soil chemistry	Chemical processes in the soil	Soil fertility, cation exchange capacity (CEC), soil pH
Soil physical and hydraulic properties	Physical and hydraulic conditions of the soil which influence the rate of water retention and flow, the presence of soil fauna and nutrient contents	Soil texture, porosity, sediment, water retention capacity, particle size composition, soil bulk density and soil compaction, soil erosion, weathering, soil loss, sediment, root depth, root density, erosion protection, soil stability, sand stability, root depth, root density, soil erodibility, soil floor

Soil biomass	Organic material belowground from living and dead plants and fungi, including roots, foliage (litter), and mycorrhiza	Soil microbial biomass, soil fungal biomass, soil organic matter (SOM), fungi to bacteria ratio, belowground productivity, root development, fine root volume and biomass
Decomposition	Process of physical and chemical breakdown of dead organic material by soil fauna (animals, fungi, microorganisms)	Decomposition rates, litter decomposition, rot

Appendix II

Table A3: RepOrting standards for Systematic Evidence Syntheses (ROSES) pro forma checklist (N.R. Haddaway et al., 2018). The table indicates, which element of the review is addressed by the standard

Section/sub-section	Topic	Description	Further explanation	Checklist/ meta-data	SR/SM
Title	Title	The title must indicate that it is either a systematic review or systematic map, and should indicate if it is an update/amendment: e.g. "…A systematic map update."	The title should normally be the same or very similar to the review question.	Meta-data	SR and SM
Type of review	Type of review	Select one of the following types of review: systematic review, systematic review update, systematic review amendment, systematic review from a systematic map, systematic map, systematic map update, systematic map amendment	See CEE Guidance on systematic mapping [1], and on amendments and updates [2]	Meta-data	SR and SM
Authors' contacts	Authors' contacts	The full names, institutional addresses and email addresses for all authors must be provided.		Checklist	SR and SM
Abstract	Structured summary	The abstract of the manuscript must not exceed 500 words and must be structured into separate sections: **Background**, the context and purpose of the review, including the review question; **Methods**, how the review was performed and statistical tests used (specifically mention search strategy, inclusion criteria, critical appraisal, data extraction and synthesis); **Results**, the main findings, including results of search and assessment of evidence base; **Conclusions**, brief summary and potential implications for policy/management and research.		Checklist	SR and SM

Background	Background	Describe the rationale for the review in the context of what is already known. Reviews must indicate why this study was necessary and what it aims to contribute to the field.	A theory of change and/or conceptual model should be presented that links the intervention or exposure to the outcome.	Checklist	SR and SM
Stakeholder engagement	Stakeholder engagement	The actual role of stakeholders throughout the review process (e.g. in the formulation of the question) must be described and explained (using a broad definition of 'stakeholder', including e.g. researchers, funders and other decision-makers; see [3])		Checklist	SR and SM
Objective of the review	Objective	Describe the primary question and secondary questions (when applicable).	The primary question is the main question of the review. The secondary questions are usually linked to sources of heterogeneity (effect modifiers).	Checklist	SR and SM
	Definition of the question components	Provide reference to the question key elements, e.g. population(s), intervention(s)/exposure(s), comparator(s), and outcome(s).	For other question types see [4,5]	Meta-data	SR and SM
Methods	Protocol	Provide citation, DOI or open-access link to published protocol.	The protocol should be peer-reviewed and publicly available online (open access).	Meta-data	SR and SM
	Deviations from protocol	Describe any ways in which the final methods of the review deviate from those set out in the protocol along with a justification.		Checklist	SR and SM

Searches	Search strategy	Detail the search strategy used, including: database names accessed, dates of searching, institutional subscriptions (or date ranges subscribed for each database), search options (e.g. 'topic words' or 'full text' search facility), efforts to source grey literature, other sources of evidence (e.g. hand searching, calls for evidence/submission of evidence by stakeholders).		Checklist	SR and SM
	Search string	Provide Boolean-style full search string and state the platform for which the string is formatted (e.g. Web of Science format)		Meta-data	SR and SM
	Languages - bibliographic databases	List languages used in bibliographic database searches		Meta-data	SR and SM
	Languages – grey literature	List languages used in organisational website searches and web-based search engines		Meta-data	SR and SM
	Bibliographic databases	Provide the number of bibliographic databases searched		Meta-data	SR and SM
	Web-based search engines	Provide the number of web-based search engines searched		Meta-data	SR and SM
	Organisational websites	Provide the number of organisational websites searched		Meta-data	SR and SM
	Estimating comprehensiveness of the search	Describe the process by which the comprehensiveness of the search strategy was assessed (i.e. list of benchmark articles)		Checklist	SR and SM
	Search update	Describe any update to searches undertaken during the conduct of the review	Compulsory (if update performed). A search update is good practice if original searches were performed more than	Checklist	SR and SM

			two years prior to review completion.		
Article screening and study inclusion criteria	Screening strategy	Describe the methodology for screening articles/studies for relevance. Methods for consistency of screening decisions (at title, abstract, and full texts levels) checking must be described.		Checklist	SR and SM
	Inclusion criteria	Describe the inclusion criteria used to assess relevance of identified articles/studies. These must be broken down into the question key elements (e.g. relevant subject(s), intervention(s)/exposure(s), comparator(s), outcome(s), study design(s)) and any other restrictions (e.g. date ranges or languages).		Checklist	SR and SM
Critical appraisal	Critical appraisal strategy	Describe here the method used for critical appraisal of study validity (including assessment of individual studies and the evidence base as a whole). Describe how repeatability of critical appraisal of study validity was tested.	Compulsory (SR) / Optional (SM)	Checklist	SR and SM
	Critical appraisal used in synthesis	Describe how the information from critical appraisal was used in synthesis.	Compulsory (SR) / Optional (SM)	Checklist	SR and SM
Data extraction	Meta-data extraction and coding strategy	Describe the method for meta-data extraction and coding for studies, providing lists of variables that will be extracted as meta-data and those that will be coded. Describe how repeatability of meta-data/data extraction and coding was tested.	Optional (SR) / Compulsory (SM)	Checklist	SR and SM
	Data extraction strategy	Describe the method for extraction of qualitative and/or quantitative study findings. Describe how repeatability of data extraction was tested.		Checklist	SR
	Approaches to missing data	Describe any process for obtaining and confirming missing or unclear information or data from authors.		Checklist	SR and SM
Potential effect modifiers/reasons for heterogeneity	Potential effect modifiers/reasons for heterogeneity	Provide a list of and justification for the effect modifiers/reasons for heterogeneity that will be considered in the review. Also provide		Checklist	SR

		details of how the list was compiled (including consultation of external experts).			
Data synthesis and presentation	Type of synthesis	State the type of synthesis conducted as part of the systematic map (narrative only) or systematic review (narrative only, narrative and quantitative, narrative and qualitative, narrative, qualitative and quantitative, narrative and mixed-methods)		Meta-data	SR and SM
	Narrative synthesis strategy	Describe methods used for narratively synthesising the evidence base in the form of descriptive statistics, tables (including SM database) and figures. Study findings must only be narratively synthesised within a SR and vote-counting must be avoided.		Checklist	SR and SM
	Quantitative synthesis strategy	If data are appropriate for quantitative synthesis, describe methods for calculating effect sizes, methods for handling complex data, statistical methods for combining data from individual studies, and any exploration of heterogeneity and publication bias. If all studies were not selected for synthesis explain criteria for selection (e.g. incomplete or missing information).	Compulsory (if quantitative synthesis performed)	Checklist	SR
	Qualitative synthesis strategy	Describe methods used for synthesising qualitative data and justify your methodological choices. Describe if and how you plan to analyse subgroups/subsets of data. If all studies were not selected for synthesis explain criteria for selection (e.g. incomplete or missing information).	Compulsory (if qualitative synthesis performed)	Checklist	SR
	Other synthesis strategies	Describe any other approaches used for synthesising data or combining qualitative and quantitative syntheses (e.g. mixed methods) and justify your choice of methodology.	Compulsory (if other synthesis performed)	Checklist	SR
	Assessment of risk of publication bias	Describe methods for examining the possible influence of publication bias on the synthesis.	This may be done for quantitative syntheses using diagnostic plots or statistical tests	Checklist	SR

	Knowledge gap and cluster identification strategy	Describe the methods used to identify and/or prioritise key knowledge gaps (unrepresented or underrepresented subtopics that warrant further primary research) and knowledge clusters (well-represented subtopics that are amenable to full synthesis via systematic review).		Checklist	SR and SM
	Demonstrating procedural independence	Describe the role of systematic reviewers (who have also authored articles to be considered within the review) in decisions regarding inclusion or critical appraisal of their own work.	Reviewers who have authored articles to be considered within the review should be prevented from unduly influencing inclusion decisions, for example by delegating tasks appropriately.	Checklist	SR and SM
Results	Description of review process	Describe the review process including the volume of evidence identified from all sources and retained through each stage of the review. Must also display the number of articles/studies included at all stages of the review in a flow diagram, including the number of articles/studies excluded at each stage.		Checklist	SR and SM
	Number of search results	Provide the number of search results from bibliographic databases (including updates if conducted) prior to duplicate removal.	This number should not include web-based search engine or organisational website searches: this will help assessment of the efficiency of the primary search string.	Meta-data	SR and SM
	Number of search results (-duplicates)	Provide the total number of search results from bibliographic database searches following duplicate removal.	This number should not include web-based search engine	Meta-data	SR and SM

		or organisational website searches: this will help assessment of the efficiency of the primary search string.		
Full text screening excludes	Additional file containing list of and reasons for full text exclusions.		Checklist	SR and SM
Title screening results	Provide the number of articles retained following title screening.	Optional if screening titles and abstracts together	Meta-data	SR and SM
Abstract screening results	Provide the number of articles retained following abstract screening.	Optional if screening titles and abstracts together	Meta-data	SR and SM
Title and abstract screening results	Provide the number of articles retained following title and abstract screening.	Optional if screening titles and abstracts separately	Meta-data	SR and SM
Retrieval results	Provide the number of articles retrieved at full text.		Meta-data	SR and SM
Unobtainable articles	Additional file containing list of unobtainable articles.		Checklist	SR and SM
Full text screening results	Provide the number of articles retained following full text screening.		Meta-data	SR and SM
Consistency checking: screening	Results of consistency checking at all stages (screening, data extraction, critical appraisal) must be provided. Provide the number of titles, abstracts and full texts screened and checked for consistency by two or more reviewers as a fraction of the total (e.g. Title: 2000/20000; Abstract: 500/5000: Full text: 10/100).		Checklist	SR and SM

	Critical appraisal exclusions	If any studies are excluded due to low validity, provide the number of studies excluded from further synthesis during critical appraisal.	Compulsory for SR if any studies not included in synthesis due to validity. Studies typically not excluded during optional critical appraisal in SM. Reviews authors may prefer to perform a sensitivity analysis (repeating analyses to examine the influence of validity) rather than excluding studies from synthesis.	Meta-data	SR
	Narrative synthesis	Describe the body of evidence identified using figures and tables, avoiding vote-counting (tallying of studies based on results; direction or significance). Each must be presented with descriptive information (meta-data) and extracted study findings (for SR). Describe the validity of individual studies and the evidence base as a whole (for SR and also for SM if critical appraisal conducted).		Checklist	SR and SM
	Extracted data	Additional file containing extracted quantitative or qualitative data (study findings) from included studies.		Checklist	SR
	Systematic map database	Additional file containing meta-data and coding for included studies.	Compulsory (SM), Optional (SR)	Checklist	SR and SM
	Quantitative synthesis	Present results of quantitative synthesis of study findings (e.g. meta-analysis).	Compulsory (if quantitative synthesis performed)	Checklist	SR

	Qualitative synthesis	Present results of qualitative analysis of study findings (e.g. summaries of identified themes or categories). Also provide additional file with the identified themes or categories for each study.	Compulsory (if qualitative synthesis performed)	Checklist	SR
	Other synthesis	Present results of any other synthesis methods used.	Compulsory (if other synthesis performed)	Checklist	SR
	Risk of publication bias	Describe the results of assessments for the possible influence of publication bias on the synthesis.	This may be done for quantitative syntheses using diagnostic plots or statistical tests	Checklist	SR
Discussion	Knowledge gaps and clusters	Describe knowledge gaps (unrepresented or underrepresented subtopics that warrant further primary research) and knowledge clusters (well-represented subtopics that are amenable to full synthesis via systematic review)		Checklist	SR and SM
	Limitations of the review	Discuss possible limitations in the methods used.		Checklist	SR and SM
	Limitations of the evidence base	Discuss possible limitations in the evidence base.		Checklist	SR and SM
Conclusions	Implications for policy/management	Summarise the state of the evidence base and discuss the way in which the identified evidence may inform policy/practice decision making in relation to the review/map question. For SR, provide any measure of the uncertainty surrounding the outcome.	Reviews must not include practical environmental management recommendations or advocacy.	Checklist	SR and SM
	Implications for research	Discuss the way in which the identified evidence may inform research including options for increasing the reliability of study design that could improve future research.	In this section some advocacy for future research on the reviewed topic is	Checklist	SR and SM

			permissible provided it is clearly justified by the review outcome/critical appraisal of study validity.		
	Identified knowledge gaps	List key knowledge gaps (unrepresented or underrepresented subtopics that warrant additional primary research).		Checklist	SR and SM
	Identified knowledge clusters	List key knowledge clusters (well-represented subtopics that are amenable to full synthesis via a systematic review).		Checklist	SR and SM
Declarations	Competing interests	Describe of any financial or non-financial competing interests that the review authors may have. Describe how the systematic reviewers who also authored articles considered within the review were prevented from influencing decisions regarding inclusion or critical appraisal of their own work.		Checklist	SR and SM

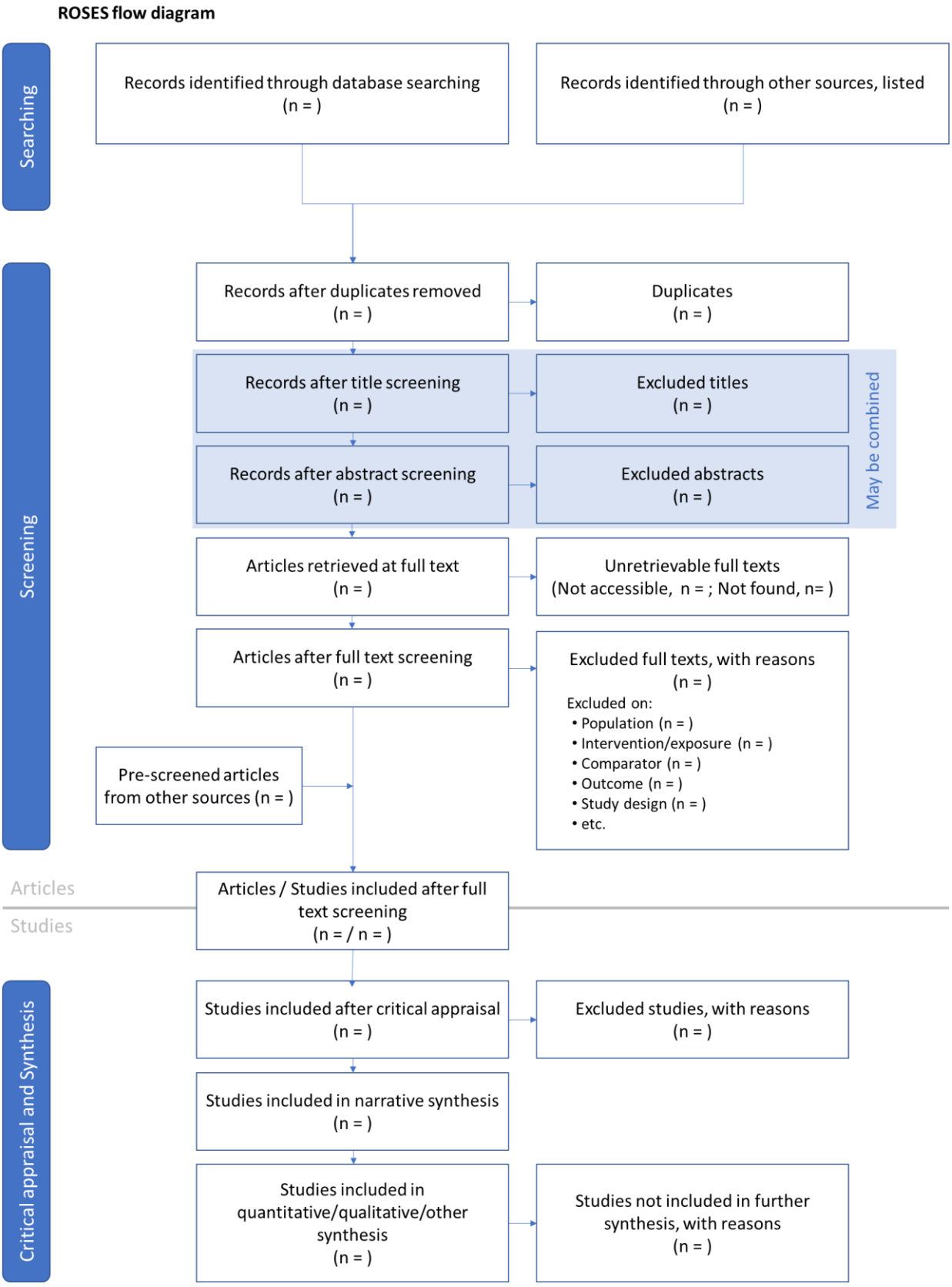

Figure A1: RepOrting standards for Systematic Evidence Syntheses (ROSES) flow diagram (N.R. Haddaway et al., 2018).

Appendix III

Table A4: Complete search strings and search details for Web of Science.

Web of Science	
SEARCH STRING	TS=((meta-analys* OR "meta analys*" OR "systematic review" OR “systematic literature review” OR (synthes* NEAR/5 (evidence OR research OR quantitative))) AND (forest* OR (plantation NEAR/5 (forest* OR tree OR timber OR stand*)) OR (stand NEAR/5 (forest* OR tree OR timber)) OR (soil NEAR/5 (forest* OR tree OR timber OR stand*))) AND (silvicultur* OR forestry OR “managed forest*” OR harvest* OR (management NEAR/5 (forest* OR tree OR timber OR stand*))))
SEARCH DETAILS	• Advanced search • Search in Web of Science Core Collection (all editions) • Exact search • The field tag *TS (TOPIC)* allows a search in title, abstract, author keywords, and Keywords Plus

Table A5: Complete search strings and search details for SCOPUS.

SCOPUS	
SEARCH STRING	TITLE-ABS-KEY ((meta-analys* OR "systematic review" OR "systematic literature review" OR (synthes* W/5 (evidence OR research OR quantitative))) AND (forest* OR (plantation W/5 (forest* OR tree OR timber OR stand*)) OR (stand W/5 (forest* OR tree OR timber)) OR (soil W/5 (forest* OR tree OR timber OR stand*))) AND (silvicultur* OR forestry OR "managed forest*" OR harvest* OR (management W/5 (forest* OR tree OR timber OR stand*))))
SEARCH DETAILS	• Advanced search • The field tag TITLE-ABS-KEY allows a search in title, abstract and keywords

Table A6: Complete search strings and search details for CAB Forest Science.

CAB Forest Science Database	
SEARCH STRING	(meta-analys* OR "meta analys*" OR "systematic review" OR “systematic literature review” OR (synthes* NEAR/5 (evidence OR research OR quantitative))) AND (forest* OR (plantation NEAR/5 (forest* OR tree OR timber OR stand*)) OR (stand NEAR/5 (forest* OR tree OR timber)) OR (soil NEAR/5 (forest* OR tree OR timber OR stand*))) AND (silvicultur* OR forestry OR “managed forest*” OR harvest* OR (management NEAR/5 (forest* OR tree OR timber OR stand*)))
SEARCH DETAILS	• Advanced search • Search in ’all fields’ • Search in *Abstracts Records*: Indexed records from the CAB Direct database relating to the subject of forest science and associated subjects

Table A7: Complete search strings and search details for Google Scholar.

Google Scholar	
SEARCH STRING	(meta-analysis OR "systematic review" OR (synthesis AROUND(5)(evidence OR research))AND(forest OR(plantation AROUND(5)(tree OR timber)) OR (stand AROUND(5)(forest OR timber)))AND(silviculture OR forestry OR(management AROUND(5)(forest OR timber OR stand)))
SEARCH DETAILS	• Advanced search - Find articles with all of the words: <SEARCH STRING ENTERED> - Where my words will occur anywhere in the article • No limitations of time, journals or authors • Search language: English • Include citations, do not include patents

Appendix IV

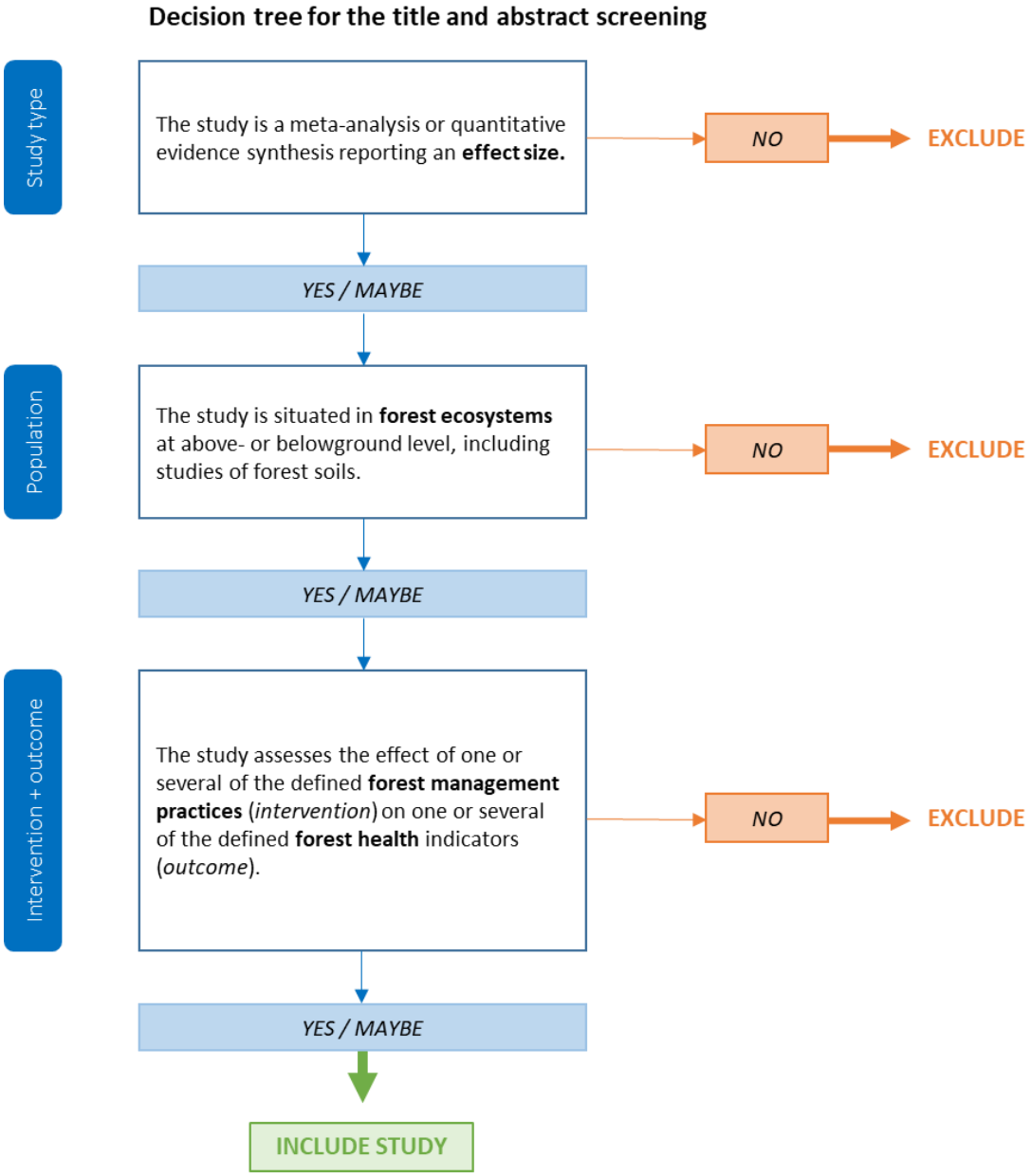

Figure A2: Decision tree for the inclusion of studies in the title and abstract screening phase. As outlined in the chapter 'In- and exclusion criteria', studies only have to fulfil the criteria A, B and C in the title and abstract screening, to be included into the full-text screening. If uncertainty remains on the fulfilment of the three criteria, the studies are also included into the full-text screening (hence, the indication 'YES/MAYBE').

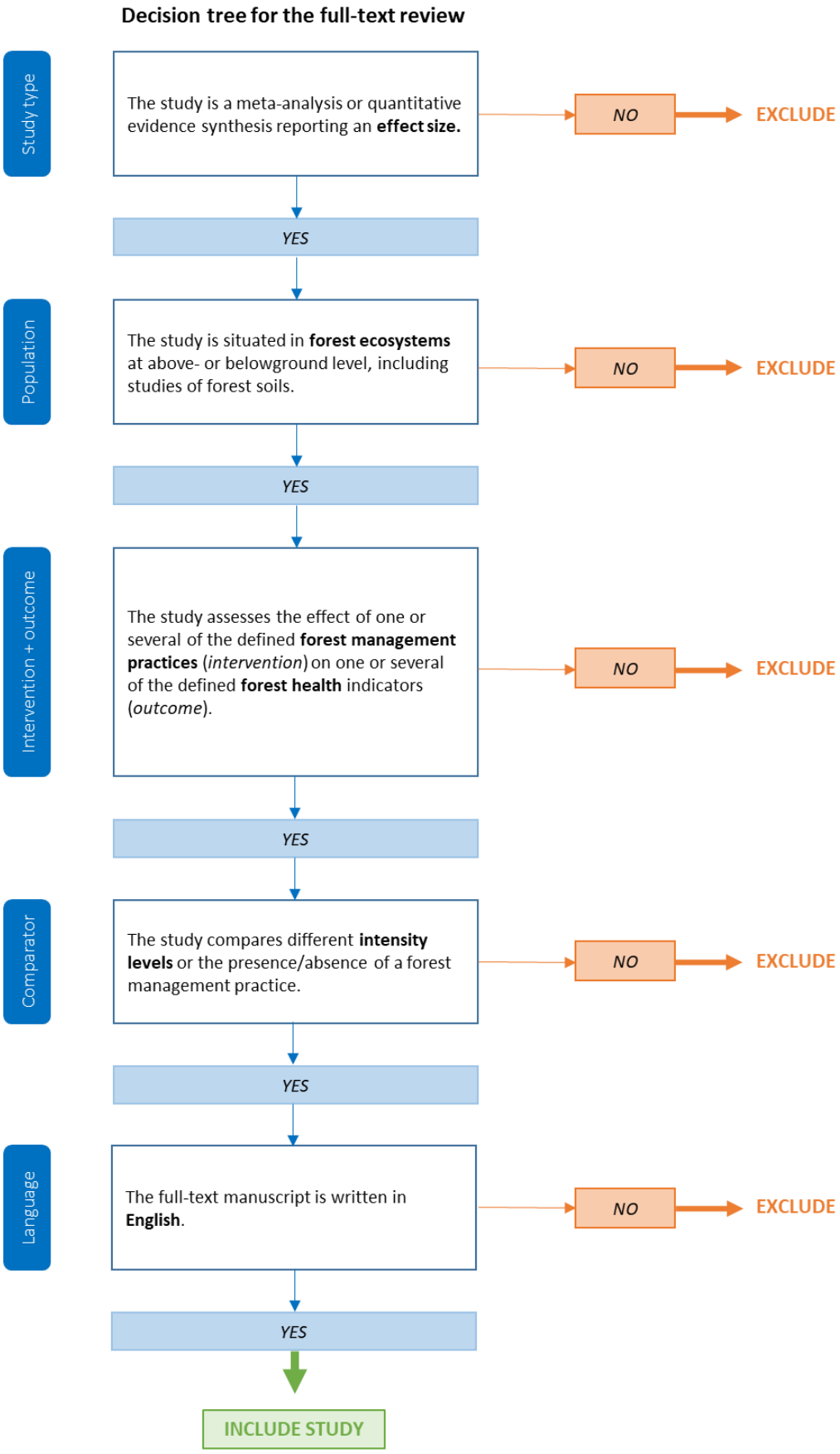

Figure A3: Decision tree for the inclusion of studies in the title and abstract screening phase. In this stage, studies have to fulfil all inclusion criteria (A-E) to be included into the umbrella review.

Appendix V

Collaboration for Environmental Evidence Synthesis Appraisal Tool (CEESAT) checklist containing 16 questions for evidence overviews (Collaboration for Environmental Evidence, 2020):

The CEESAT Checklist

1. THE REVIEW QUESTION
    1.1. Are the elements of the review question clear?
2. THE METHOD/PROTOCOL
    2.1. Is there an a-priori method/protocol document?
3. SEARCHING FOR STUDIES
    3.1. Is the approach to searching clearly defined, systematic and transparent?
    3.2. Is the search comprehensive? [3]
4. INCLUDING STUDIES
    4.1. Are eligibility criteria clearly defined?
    4.2. Are eligibility criteria consistently applied to all potentially relevant articles and studies found during the search?
    4.3. Are eligibility decisions transparently reported?
5. CRITICAL APPRAISAL
    5.1. Does the review critically appraise each study?
    5.2. During critical appraisal was an effort made to minimise subjectivity?
6. DATA EXTRACTION
    6.1. Is the method of data extraction fully documented?
    6.2. Are the extracted data reported for each study?
    6.3. Were extracted data cross checked by more than one reviewer?
7. DATA SYNTHESIS
    7.1. Is the choice of synthesis approach appropriate?
    7.2. Is a statistical estimate of pooled effect (or similar) provided together with measure of variance and heterogeneity among studies?
    7.3. Is variability in the study findings investigated and discussed?
8. LIMITATIONS
    8.1. Have the authors considered limitations of the synthesis?

[3] As stated in the critical appraisal sub-chapter, the criterion 3.2 was also scored RED, if Google Scholar was used as the sole search engine.

Appendix VI

Table A8: AACODS checklist for the evaluation and critical appraisal of the quality of grey literature (Tyndall, 2010).

AACODS element	Question	YES	NO	MAYBE
Authority	Identifying who is responsible for the intellectual content. **Individual author:** • Associated with a reputable organisation? • Professional qualifications or considerable experience? • Produced/published other work (grey/black) in the field? • Recognised expert, identified in other sources? • Cited by others? (use Google Scholar as a quick check) • Higher degree student under "expert" supervision? **Organisation or group**: • Is the organisation reputable? (e.g. W.H.O) • Is the organisation an authority in the field? **In all cases:** • Does the item have a detailed reference list or bibliography?			
Accuracy	• Does the item have a clearly stated aim or brief? • Is so, is this met? • Does it have a stated methodology? • If so, is it adhered to? • Has it been peer-reviewed? • Has it been edited by a reputable authority? • Supported by authoritative, documented references or credible sources? • Is it representative of work in the field? • If No, is it a valid counterbalance? • Is any data collection explicit and appropriate for the research? • If item is secondary material (e.g. a policy brief of a technical report) refer to the original. Is it an accurate, unbiased interpretation or analysis?			
Coverage	All items have parameters which define their content coverage. These limits might mean that a work refers to a particular population group, or that it excluded certain types of publication. A report could be designed to answer a particular question, or be based on statistics from a particular survey. • Are any limits clearly stated?			
Objectivity	It is important to identify bias, particularly if it is unstated or unacknowledged. • Opinion, expert or otherwise, is still opinion: is the author's standpoint clear? • Does the work seem to be balanced in presentation?			

Date	For the item to inform your research, it needs to have a date that confirms relevance • Does the item have a clearly stated date related to content? No easily discernible date is a strong concern. • If no date is given, but can be closely ascertained, is there a valid reason for its absence? • Check the bibliography: have key contemporary material been included?			
Significance	This is a value judgment of the item, in the context of the relevant research area • Is the item meaningful? (this incorporates feasibility, utility and relevance) • Does it add context? • Does it enrich or add something unique to the research? • Does it strengthen or refute a current position? • Would the research area be lesser without it? • Is it integral, representative, typical? • Does it have impact? (in the sense of influencing the work or behaviour of others)			