Thesis

presented by

Maria Adelia WIDIJANTO

in partial fulfillment of the requirements
for the Degree of Master of Science
in Agriculture and Environmental Sciences (ST2AE)
track: Global Forestry

Title

Evaluating Natural Climate Solution Metrics in the Tropics

Publicly defended on 05/10/2025

at AgroParisTech, France

in front of the following jury:

Dr Sylvain SCHMITT	Examiner
Dr Raphaël MANLAY	Academic supervisor 1
Dr Henri BAZZI	Academic supervisor 2
Ping YOWARGANA	Host supervisor

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the Global Forestry EMJM Committees, particularly the faculty and staff at the University of Copenhagen, TU Dresden, and AgroParisTech, for granting me the opportunity to expand my understanding of forests around the world. Coming from a country blessed with vast natural resources, this program has allowed me to view my homeland's forests from multiple academic and cultural perspectives. The courses have been highly influential in shaping the way I approached and developed this master's thesis.

I am sincerely grateful to the Erasmus Mundus Joint Masters Programme for the generous scholarship support that made it possible for me to pursue this two-year journey across multiple countries and universities. I also wish to extend my gratitude to IIASA for funding and hosting my internship. The experience not only strengthened the foundation of my thesis but also gave me exposure to international discussions on environment and climate policy.

I would also like to express my gratitude to my academic supervisors, Dr. Raphaël Manlay and Dr. Henri Hassan Bazzi, for their guidance and constructive feedback throughout the development of this thesis. I am equally thankful to my host supervisor, Ping Yowargana, for his constant encouragement to strive beyond my limits and for inspiring me to grow not only as a researcher, but also as a person.

Above all, I owe my deepest gratitude to my family for their steady support and encouragement throughout the course of this master's program. Their understanding and belief in my pursuit of this academic path have been an important source of strength, allowing me to stay focused and committed to completing this thesis.

Lastly, I would like to thank my friends, whose kindness and companionship made the journey of completing this thesis far lighter. To those who welcomed me into their homes in Vienna, Aarhus, and Jakarta, offering both space and support, and to those who were always just a call away with encouragement and reminders to keep moving forward – I am deeply grateful. I also want to express my heartfelt thanks to my GLOFOR cohort friends, whose solidarity, humor, and shared determination made even the most challenging moments feel manageable. Without your constant support, gentle nudges, and occasional well-timed distractions, this thesis might still be a draft.

ABSTRACT

The land sector plays a critical role in climate change mitigation, with tropical regions having most of the global potential area for nature climate solutions (NCS). However, assessing the effectiveness of NCS remains challenging, as it requires appropriate metrics that can accurately capture the changes these interventions produce, yet such metrics are often limited in their adoption due to practical constraints. To better understand these limitations and how they shape metric usage, this study systematically maps metrics that have been applied in the literature to measure NCS interventions in Brazil, Indonesia, and the Democratic Republic of Congo, which represent most pantropical NCS potential. This study specifically analyzed both the methodological complexity and operational requirements of these metrics to examine how these factors influence their adoption in practice. The systematic mapping identified 235 relevant studies where 116 unique metrics were used. Biomass-related metrics are the most frequently used metrics, while metrics related to human wellbeing were notably underrepresented. Metrics with higher methodological complexity tended to be used in a greater number of studies, especially those with established protocols. Operational requirements varied across metrics, where metrics such as soil properties and landscape connectivity were commonly associated with technology-intensive approaches such as laboratory analysis and remote sensing. In contrast, biodiversity and biomass metrics relied heavily on labor-intensive fieldwork and social data collection. Finally, four archetypes emerged from the joint analysis of complexity and operational requirements, offering a framework to guide resource allocation and future metric development.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

Since the Kyoto Protocol in 1997 formally recognized its important role in climate change mitigation, the agriculture, forestry, and other land use (AFOLU) sector has remained a central arena for mitigation efforts. Over the past two decades, international climate governance has expanded to include a range of AFOLU-related interventions, such as the reducing emissions from deforestation and forest degradation in developing countries program (REDD+) (Wunder *et al.*, 2020) or ecological engineering (Mitsch, 2012). Alongside this expanding knowledge of AFOLU interventions, the concept of natural climate solutions (NCS) emerged with clearer boundaries of a climate intervention (Griscom *et al.*, 2017). Defined by principles such as being nature-based, climate-additional, measurable, and equitable, NCS serve as a unifying framework to ensure interventions deliver both climate benefits and human wellbeing outcomes (Ellis *et al.*, 2024).

Although these principles provide a conceptual basis, in practice, NCS interventions raise several challenges. NCS interventions involve complex ecosystems, where active management can change the interactions of the ecosystem's components (Maes *et al.*, 2024). Measuring each of the components' changes using limited metrics can lead to underestimation or overestimation of the broader ecosystem benefits. Acknowledging a wider range of metrics as equally valid for comparison can prevent the loss of nuance, especially to quantifying the co-benefits of NCS interventions (Ellis *et al.*, 2024 ; Chang *et al.*, 2024).

The tropical region holds the highest potential area for NCS, contributing to 61% of the total global NCS potential (Griscom *et al.*, 2017). Humans play a crucial role in managing tropical ecosystems, and human activities influence many environmental issues in these regions (Pyles *et al.*, 2022). Community and stakeholder engagement in tropical landscapes is therefore essential, as their perceptions of NCS interventions significantly affect both the success and sustainability of these efforts. (Nerfa *et al.*, 2021 ; Höhl *et al.*, 2020). Stakeholders are generally aware of their internal institutional constraints in conducting NCS interventions; however, they often lack the necessary insights on how to manage these limitations effectively (Brumberg *et al.*, 2025), such as how limited resources should be allocated to gather foundational information for planning, monitoring, and evaluation of the interventions (Mahajan *et al.*, 2023). This includes the choosing of metrics: failure to apply the appropriate metrics for monitoring and evaluation of NCS interventions can result in incomplete assessment, which could ultimately lead to misalignment with project goals and wasted resources (Sowińska-Świerkosz & García, 2021).

Previous studies have developed frameworks to guide the design of monitoring and evaluation systems, mainly for nature-based solutions projects (Rödl & Arlati, 2022 ; Sowińska-Świerkosz & García, 2021), including land restoration activities (Prach *et al.*, 2019 ; Convertino *et al.*, 2013 ; Morgan *et al.*, 2022). These studies highlight that setting clear targets is the first step in identifying appropriate metrics. However, applying these metrics in practice requires practical considerations beyond the substantial alignment of the project's goal. This study focuses on two dimensions of the practical considerations: the complexity of existing available methodologies to measure the metric, and the operational requirements when implementing those metrics.

Methodological complexity

NCS intervention introduces changes in a certain aspect of an ecosystem component (e.g., an intervention can increase or decrease biodiversity). These changes can be represented by a unit of metric (e.g., species richness). However, ecosystem components are inherently complex, and different measurement techniques may capture different information about the same metric (e.g., measuring species richness using visual surveys may underestimate cryptic or nocturnal species compared to measurement using acoustic or environmental DNA methods). Ecosystems can be viewed as classic examples of complex systems, in which both stable and unpredictable processes influence their behavior (Parrott, 2010). As a result, no single measure can fully describe their nature, making the use of diverse methodologies essential for capturing their complexity.

This study uses methodological complexity to describe the extent of existing methods to measure a metric. Metrics with low methodological complexity have only a few established available methods. In contrast, highly complex metrics represent multifaceted systems with diverse measurement approaches. In this varying options of methods to measure a metric, using a method that is not suitable for the context can lead to a misleading conclusion (Rödl & Arlati, 2022).

Operational requirements

Operational requirements include the practical conditions and resources needed to carry out a measurement, such as equipment, personnel, time, and financial resources. Considering these requirements helps decision-makers select methods that not only accurately capture outcomes but also align with available budgets and capacities (Brumberg *et al.*, 2025). Costs extend beyond the purchase of instruments to include labor and logistical support, which are often overlooked "hidden costs" that can affect the success of NCS interventions. Acknowledging the operational requirements of a metric is important to ensure that measurements are both feasible and effective.

Understanding these two aspects is important for achieving feasible and reliable measurements that align with project goals, optimize resource allocation, and ensure accurate monitoring and evaluation of NCS interventions.

1.2. RESEARCH QUESTIONS

This study aims to answer the following questions:

1. What is the current state of metric usage for evaluating NCS interventions in the tropics?
2. What patterns exist between the methodological complexity of metrics and their use in studies evaluating NCS interventions?
3. What patterns exist between the operational requirements of metrics and their use in the studies evaluating NCS interventions?

To answer, this study uses the systematic mapping approach. Systematic mapping is a methodology to gather evidence from existing studies with a replicable, objective, and transparent process. Unlike a systematic review, which synthesizes the results of past studies to determine an overall effect, the systematic mapping approach focuses on creating a catalog of evidence. Its goal is to map the existing research efforts and identify areas where more study is needed, rather than to compile and analyze the research's findings (James et al., 2016).

2. METHODOLOGY

The Reporting Standards for Systematic Evidence Syntheses (ROSES) guideline was adopted to develop the systematic mapping protocol (Haddaway *et al.*, 2018). This methodology has been used in other systematic mapping studies in the environmental field (e.g., Chausson *et al.*, 2020 ; Paxton *et al.*, 2024 ; Westwood *et al.*, 2023).

2.1. SCOPE OF STUDY

To define clearly the scope of this study, this study's research questions are broken down into the Population, Intervention, Comparator, and Outcomes (PICO) framework (Pullin *et al.*, 2022), except for the Comparator component, as it is not relevant to be defined for this study (Table 1). This study focuses on NCS studies conducted in the tropical regions of Brazil, the Democratic Republic of Congo (DRC), and Indonesia, as they are countries that hold more than half of the cost-effective NCS potential in the pantropical region (Griscom *et al.*, 2020).

Table 1 Scope of the study using the PICO framework

PICO Component	Guiding Questions	Scope Definition
P – Population	What and where is the NCS intervention conducted?	Terrestrial tropical ecosystems in the tropical parts of Brazil, the Democratic Republic of Congo (DRC), and Indonesia
I – Intervention	What types of NCS interventions are being evaluated?	All activities align with the NCS principles (Ellis *et al.*, 2024) and pathways (Griscom *et al.*, 2017)
O – Outcome	What are the reported metrics? How are they measured?	The study's outcome must be measurable, replicable, and comparable across different studies.

2.2. DATA ACQUISITION

2.2.1. SEARCH STRATEGY

The search was conducted on the Web of Science CORE Index Collections and Scopus. Both databases' search was limited to publications using the English language. In Scopus, the search is restricted to documents classified as articles, conference papers, or reviews. In Web of Science, the search excludes content from the Preprint Citation Index and includes only documents categorized as articles, review articles, or other types.

The initial search string was developed based on the scope of the study's PICO framework, where each of the components was defined as a set of keywords. The search strings were tested and refined iteratively to balance comprehensiveness and relevance of words commonly used in publications. To support this process, *litsearchr* R package (Grames *et al.*, 2019) was used to ensure the chosen strings corresponded to the most frequent terms used in the publications' titles and keywords. This R package was used only to verify whether the suggested keywords helped reduce the number of irrelevant publications. The final search string was manually selected to ensure alignment with the study's scope (Appendix 1).

The search conducted on May 7, 2025, identified 2,888 publications (1,110 results from Scopus and 1,778 results from Web of Science). Duplicates from both databases were removed using *synthesisr* R package (Westgate & Grames, 2025). After duplicates were removed, 2,066 unique publications remained for further screening (Figure 1).

2.2.2. SCREENING PROCEDURE

The screening procedure consists of two steps: first, by the title and abstract, then by the full text. These are the exclusion criteria used conservatively in both screening steps.

1. Studies without measurable output, which cannot be replicated and compared across different studies (e.g., descriptive analysis).
2. Studies focusing on ecosystems other than the AFOLU sector.
3. Studies measuring the degree of ecosystem or the natural recovery process of an ecosystem when NCS activities are absent.
4. Studies focusing on identifying the drivers behind the intervention, rather than direct measurement of the ecosystem health change where the intervention was conducted.
5. Studies outside Indonesia, Brazil, or DRC, except for cross-country studies where these three countries are included.
6. Studies focusing on the advancement of ecosystem monitoring technologies.
7. Studies which activities did not follow NCS principles (Ellis *et al.*, 2024).
8. Review studies (e.g., systematic reviews, meta-analyses, or literature reviews).

Title and abstract screening was done using Rayyan as the supporting tool (Ouzzani *et al.*, 2016). After full-text screening, 235 publications were included in the systematic map database for further data coding and evidence synthesis analysis.

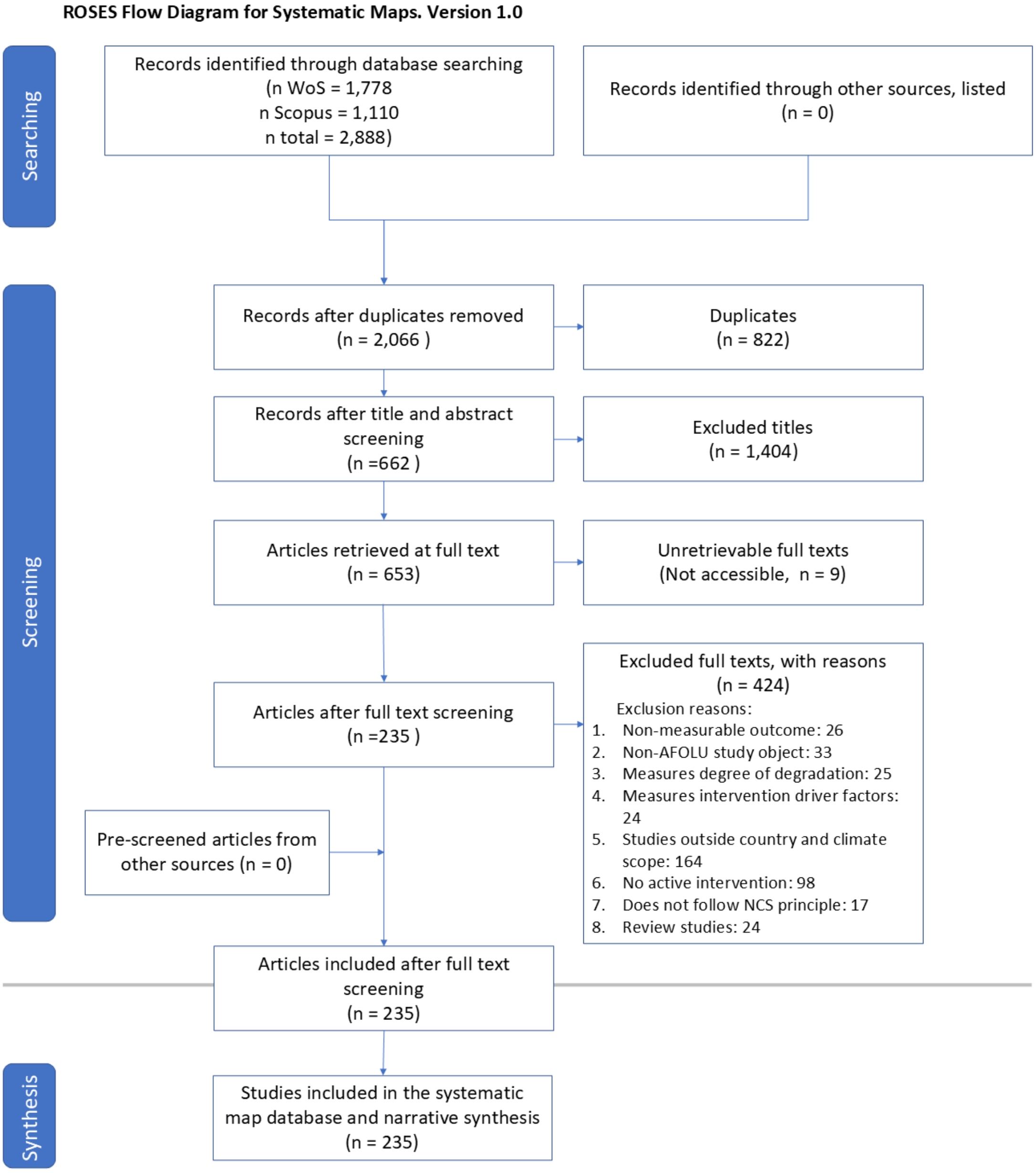

Figure 1 Flow diagram of screening procedure based on ROSES protocol guideline (Haddaway *et al.*, 2018)

2.2.3. DATA EXTRACTION AND CODING

Information from the selected publications is then extracted into a database. The database's columns correspond to the relevant information needed to answer the research question (Figure 2). This includes the study's methodology description, intervention pathway, metric used to report the study's results, and its evaluation criteria. The complete description of each of the columns is provided in the Appendix 2. Each row of the database contains unique metrics used in the studies, and one study can have more than one metric used to report its results. Google Form was used as an interface to support the data extraction process.

The methodology used in the studies was categorized into six broader method groups. This typology was adapted from the data collection method grouping used in another systematic mapping study (Key *et al.*, 2022). Each metric within a study could be measured with more than one of these method groups. Examples of actual methods used in studies are shown in Table 2.

Table 2 Examples of methodology details in each method group

Method group	Example of methods
Ground visual survey	Vegetation survey, in-situ measurement tools (e.g., soil cation tester, tensiometer)
Laboratory analysis	Dried-oven weighting of samples, gas chromatography
Modeling	Computational tools using simulation models, statistical models, GIS-based modeling
Remote sensing	Satellite imagery, drone-based surveys, LiDAR, multispectral/hyperspectral sensors
Secondary data analysis	Use of government statistics, scientific databases, published literature, archival data
Social collection	Household surveys, interviews, focus groups, participatory rural appraisal

The description of the activity conducted in the studies was classified into 11 NCS pathway options (Griscom *et al.*, 2020). The "manage" pathways include improved fire management, natural forest management, nutrient management, optimal grazing intensity, and the integration of trees within agricultural lands. The "protect" pathways include avoided forest conversion, avoided mangrove impact, and avoided peat impact. The "restore" pathways include reforestation, mangrove restoration, and peatland restoration.

The study's outcomes were assessed at two levels of granularity: metrics and evaluation criteria. At the metric level, the original outcome descriptions directly from each study were extracted and then standardized by assigning consistent terms to metrics that conveyed the same type of information about an intervention's output. These standardized metrics were then grouped into eight distinct evaluation criteria. These evaluation criteria were organized around three overarching NCS benefits (Chang *et al.*, 2024 ; Ellis *et al.*, 2024): climate (biomass, greenhouse gas), environment (biodiversity, biogeochemical properties, ecosystem structure/function/dynamics), and human wellbeing (economic, human wellbeing). Each metric can only be assigned exclusively to an evaluation criterion.

The final database was cleaned by removing duplicates, checking for missing values, and standardizing categorical entries. All subsequent analyses were carried out using R version 4.4.2.

2.3. DATA PREPARATION

2.3.1. METHODOLOGICAL COMPLEXITY SCORING

Shannon-index (H') was used to analyze the methodological complexity of each metric. This index was originally developed in information theory as an entropy concept to quantify uncertainty within a dataset (Spellerberg & Fedor, 2003 ; Shannon, 1949). This measure provides a way to quantify how varied and balanced methodological availability of a metric are by looking at both richness (number of method groups) and evenness (distribution among the use of method groups) within a metric. The H' index for each metric was measured using Equation 1.

Equation 1

$$\mathrm{H}' = -\sum_{\mathrm{i}=1}^{\mathrm{S}} P_i \cdot \log_2 P_i$$

where P_i is the proportion of method group i within a metric, and S is the total number of distinct method groups used in the study. H' value approaching zero indicates low methodological complexity, where a certain method group dominated the measurement approaches, and a value of zero occurs when only a single method group was used to measure the metric. In contrast, a higher H' indicates a metric has been measured equally with varied methods, with no certain method being used more dominantly than others. To facilitate comparison across metrics, the H' values were rescaled to a 1-5 range using min-max normalization, as shown in Equation 2.

Equation 2

$$x' = 1 + \frac{(x - \min(x)) \cdot (5 - 1)}{\max(x) - \min(x)}$$

where x is the original H' value and x' is the rescaled score.

2.3.2. OPERATIONAL REQUIREMENT SCORING

We defined operational requirements in three aspects: the measurement technology's ease of usage, the intensity of labor needed to measure the metrics in full-time equivalent (FTE), and the total estimated monetary expense in monetary units required to measure one standardized unit (e.g., one hectare, one complete measurement cycle). For each method group, operational requirement scores were assigned for three aspects: equipment, labor, and cost, each on a 1-3 scale according to predefined guidelines (the "Operational Requirement" description row in Appendix 2). This 3-scale was used to make a concise judgment of each method group's operational requirement, which allows for a simpler and more standardized scoring. The final scores for each method group are presented in Appendix 3.

These scores were then applied to the method groups used in evaluation criteria, where each of the evaluation criteria could be associated with one or more method groups identified in the studies. For each evaluation criterion, an equally weighted mean of the three operational requirement aspects (equipment, labor, and cost) was calculated. To enable direct comparison with Methodological Complexity, the resulting operational requirement scores were rescaled to a 1–5 range using Equation 2.

2.4. STATISTICAL ANALYSIS

Multiple Correspondence Analysis (MCA) was used to assess whether the grouping of metrics under evaluation criteria was supported by the underlying data structure. In this analysis, the method group, simplified NCS intervention pathway (protect, manage, restore), and land-use context were included as descriptor variables. With MCA, these multidimensional descriptor variables can be represented along a reduced set of dimensions that capture the main patterns in the data (Le Roux & Rouanet, 2010). The resulting MCA coordinates were then used as input for a Multivariate Analysis of Variance (MANOVA) with evaluation criteria as the grouping factor. The MANOVA result shows whether metrics grouped under the same evaluation criteria are significantly different based on their descriptor variables pattern. In general, MCA and MANOVA were used to confirm that the evaluation criteria groupings reflect meaningful patterns of metric characteristics through the descriptor variables, which then can be used as a valid framework for subsequent analyses.

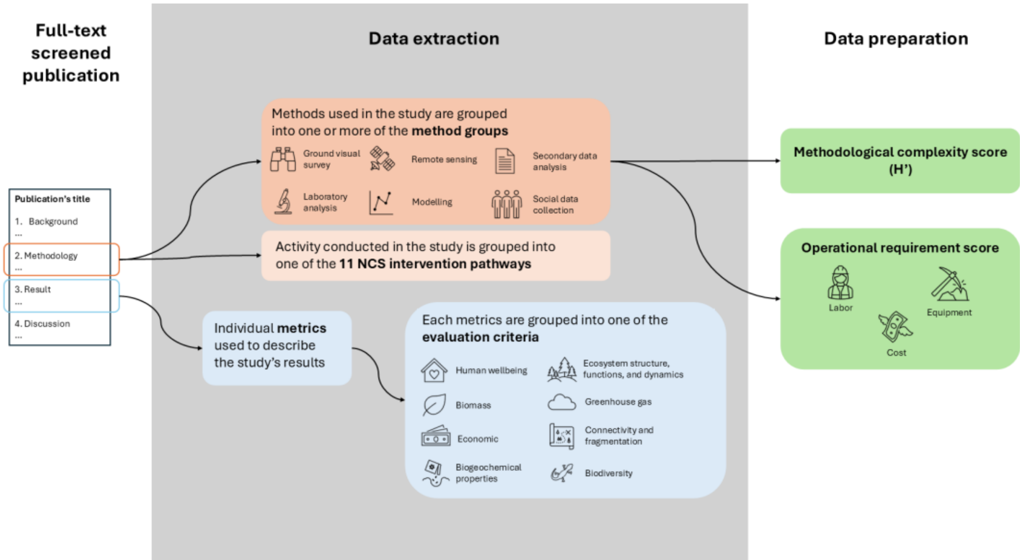

Figure 2 Conceptual diagram of the extracted variables for the analysis

3. RESULTS

3.1. OVERVIEW OF NCS METRICS USED IN STUDIES

Across the 235 studies identified, most studies were conducted in Brazil (n = 133; 56.6%), followed by Indonesia (n = 81; 34.5%). Study from DRC was found the least in the literature (n = 10; 4.3%).

After aggregating overlapping metrics used across studies, 116 unique metrics were identified (Appendix 4). MCA was performed to summarize the relationship between these metrics based on the descriptor variables, and the resulting coordinates were analyzed using MANOVA (Appendix 5). The MANOVA result on the first five dimensions of the MCA coordinates showed a significant effect of the evaluation criteria grouping of metrics (Pillai's trace = 0.43, $F(35, 1370) = 3.69$, $p < 0.001$). This result suggests that the evaluation criteria classification captures meaningful variation in metric characteristics, thereby supporting its validity as an analytical framework for the rest of this study's analysis.

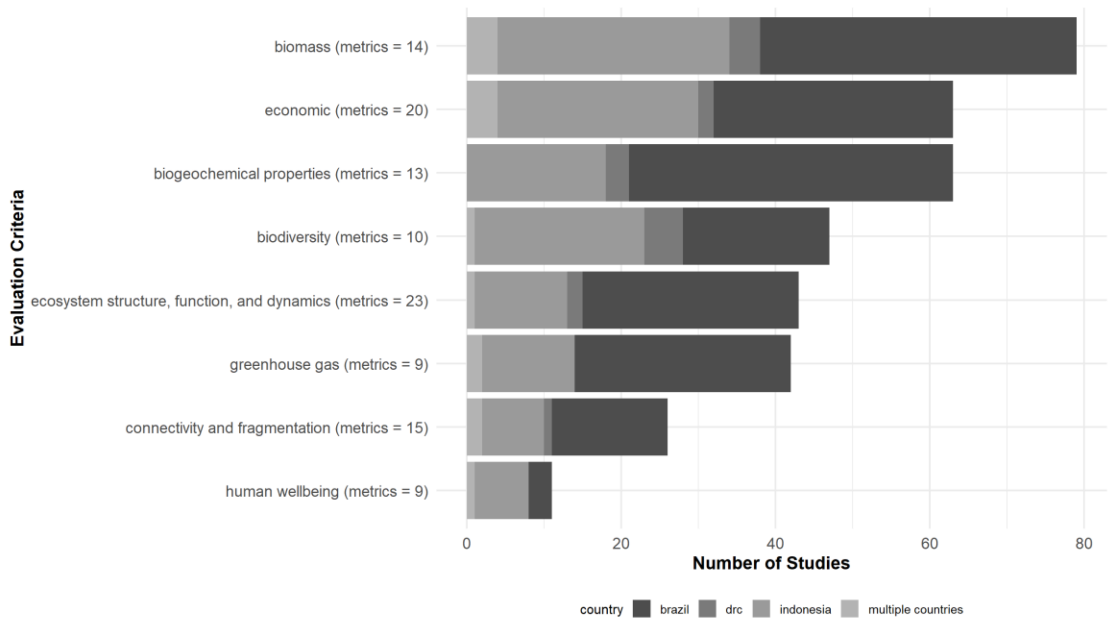

Figure 3 Number of studies using the metrics within each evaluation criteria. The number in parentheses on the Y-axis indicates how many individual metrics are included within each evaluation criterion.

Metrics within biomass evaluation criterion were used the most in the studies (n = 79), but the evaluation criterion itself includes relatively few distinct metrics (14), which indicates that those studies repeatedly used the same metrics. Ecosystem structure/function/dynamics evaluation criterion has the most variety of metrics (23), and, in contrast, human wellbeing has the lowest number of metrics (9) and is the most rarely used evaluation criterion in studies (n = 11). Most of the evaluation criteria can be found in all countries, except for metrics related to greenhouse gas and human wellbeing which were not found in the studies conducted in the DRC (Figure 3).

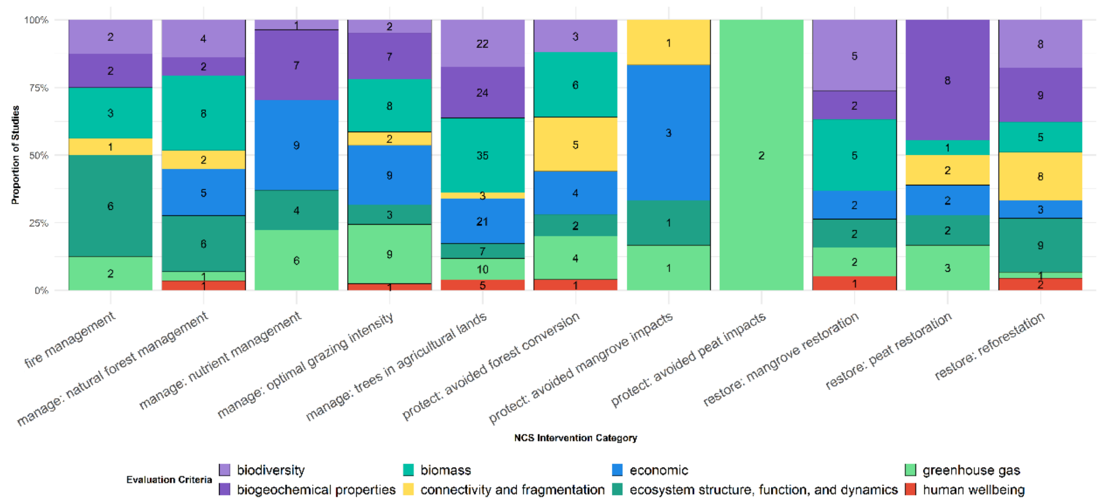

Figure 4 Proportion of evaluation criteria used in different intervention pathways. The number inside each column indicates the number of studies

There are no consistent patterns of particular evaluation criteria being applied to specific intervention pathways (Figure 4). Nevertheless, the dominant evaluation criterion within each pathway shows the aspects of the ecosystem most frequently assessed in NCS interventions. In many cases, NCS intervention pathways often measures the environmental dimensions of the ecosystem. For instance, fire management tends to focus on ecosystem structure/function/dynamics (37.5%); peat restoration emphasizes biogeochemical properties (44.4%) – largely soil-related metrics (details in Appendix 4); while mangrove restoration measures biomass and biodiversity equally (26.3%).

By contrast, some other pathways focus more on socioeconomic outcomes rather than biophysical changes. For instance, in the avoided mangrove impacts pathway, most studies assess the economic benefits of conservation (50%), while nutrient management interventions are frequently evaluated in economic terms (33.3%).

3.2. METHODOLOGICAL COMPLEXITY

The methodological complexity of metrics varied across different evaluation criteria (Figure 5). Overall, most methodological complexity scores exhibited a right-skewed distribution, indicating that the majority of metrics across these criteria used low methodological complexity method groups. In contrast, the greenhouse gas evaluation criterion was the only group with a left-skewed distribution. Although it shares the same number of metrics as human wellbeing (metrics = 9), greenhouse gas metrics show higher methodological complexity. This highlights a key difference: greenhouse gas studies employ diverse, high-complexity methods focused on a few metrics, while human wellbeing metrics rely on consistently similar, low-complexity approaches.

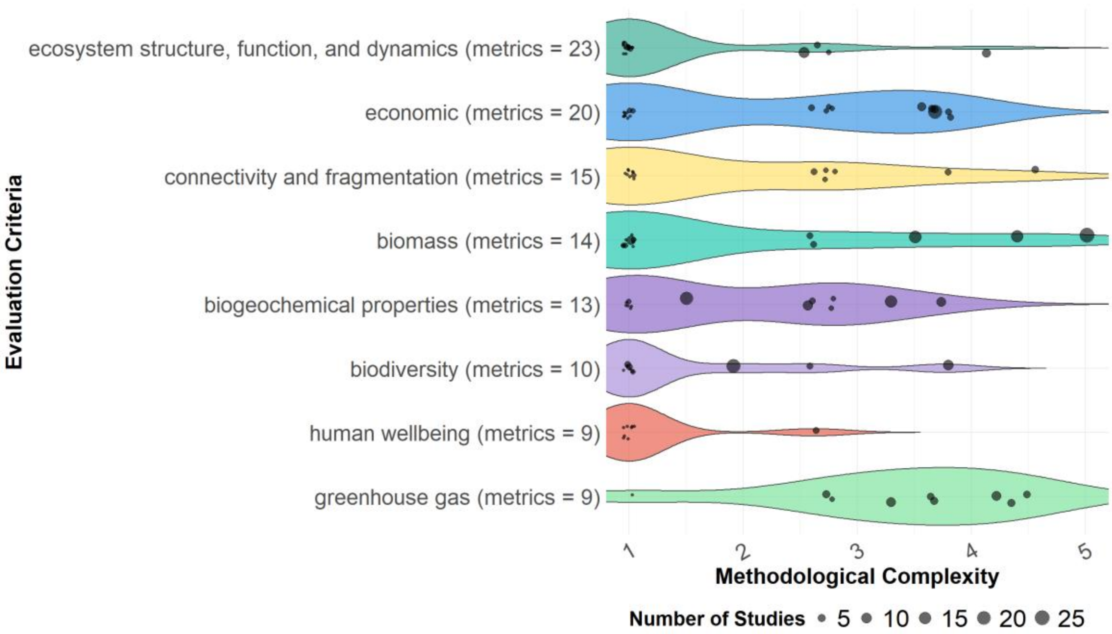

Figure 5 Methodological complexity distribution of metrics within each evaluation criterion

There is a positive correlation between methodological complexity and the number of studies per metric. Notably, metrics with high complexity scores (above the median of their evaluation criteria) are also supported by a larger body of research (defined as the top 5% of study counts). These metrics include aboveground biomass (complexity = 5; 25 studies), crop yield (3.66; 22 studies), soil chemical properties (3.27; 14 studies), soil organic carbon (4.38; 14 studies), soil physiochemical properties (1.54; 18 studies), total carbon stock (3.51; 16 studies), and vegetation species diversity (1.93; 22 studies).

3.3. Operational requirements

Figure 6 shows the method group trends of use among evaluation criteria. This is important to illustrate how the trend of the method group used by metrics within each evaluation criterion influences the evaluation criterion's operational requirements. Evaluation criteria with fewer studies, such as human wellbeing (n = 11), tend to have operational requirement scores influenced by a smaller range of methods, which may lead to less variability in their scores. In contrast, evaluation criteria supported by a larger number of studies, such as biomass (n = 79), show greater diversity in methodological approaches and thus more balanced operational requirement profiles. Ground visual survey is the most used method group among the studies (n = 133), followed by its combination with laboratory analysis (n = 90). This trend of method use for each evaluation criterion directly informs its operational requirements, with the specific scoring for each method group detailed in Appendix 2.

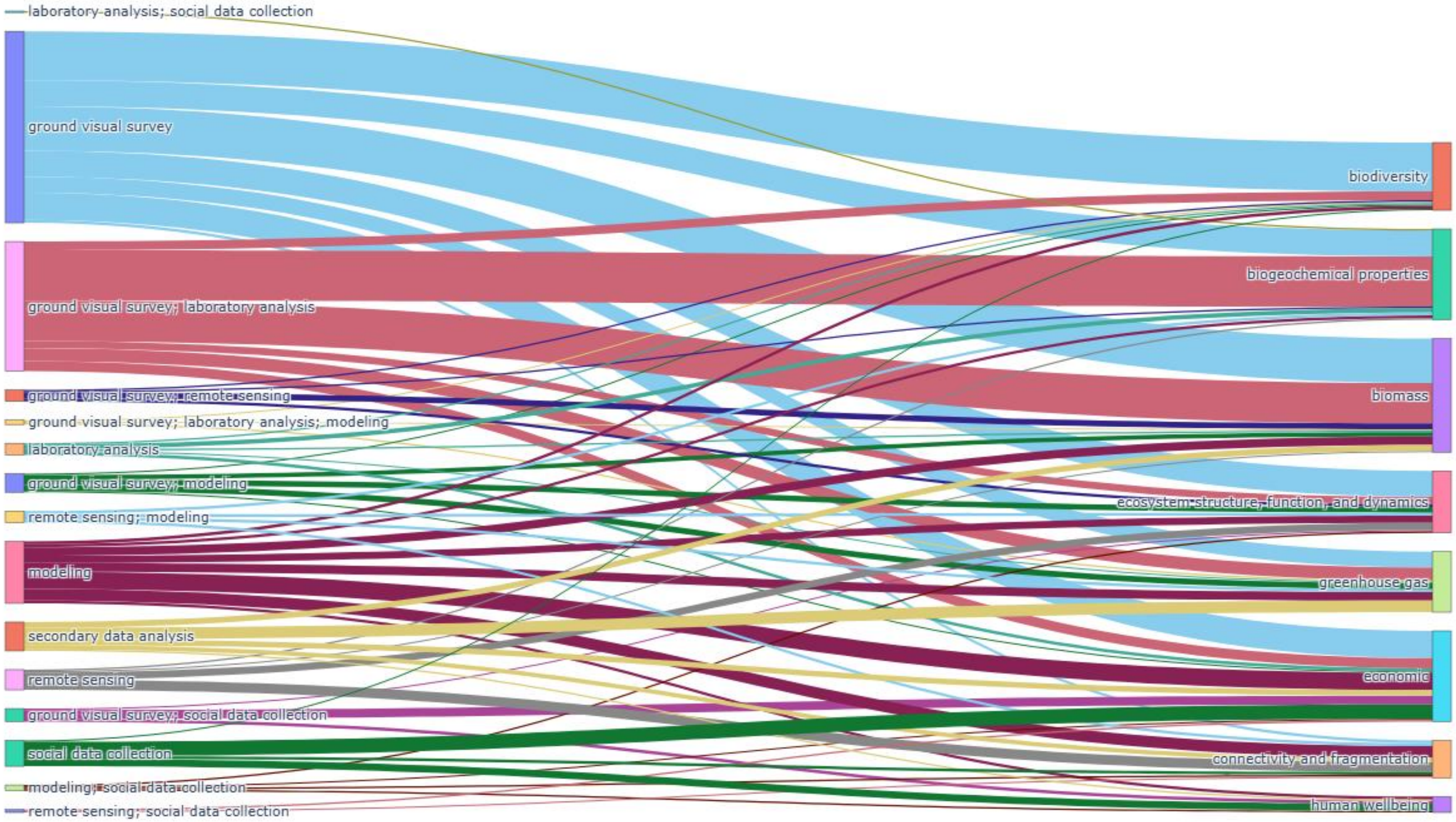

Figure 6 Method groups used in each evaluation criterion

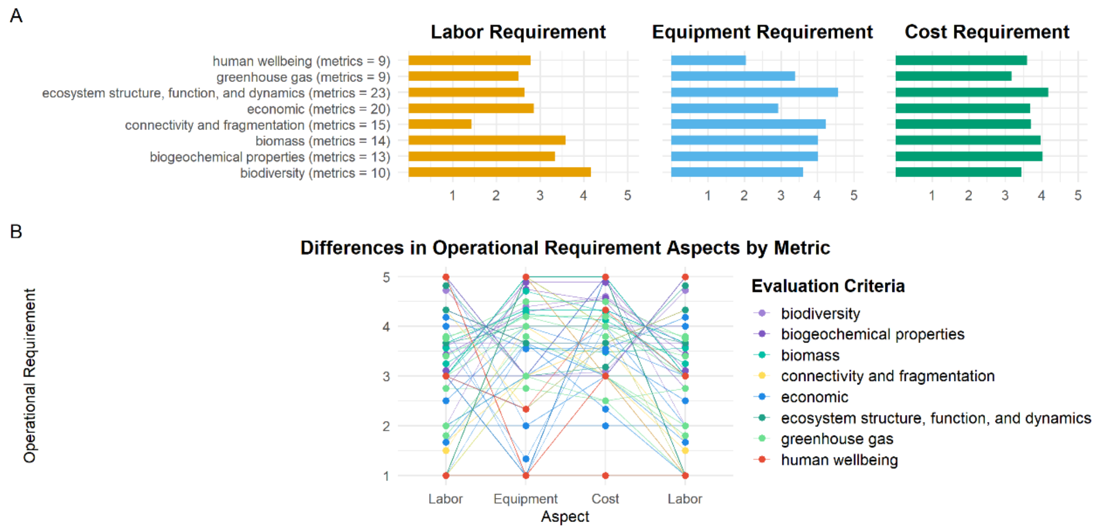

Figure 7 (A) Average equipment, labor, and cost requirements of metrics within each evaluation criterion, and (B) the distribution of individual metrics – color-coded by their evaluation criteria. When many lines slope sharply between two axes, it suggests a trade-off between two operational requirement aspects.

Figure 7A highlights the operational requirement dimensions that drive the resource use for each criterion. Ecosystem structure/function/dynamics has the highest equipment and cost requirements. This is because, as shown in Figure 6, it relies on resource-intensive methods such as laboratory analysis and modeling. On the other hand, biodiversity evaluation criterion has been heavily measured using ground visual survey, which is a labor-intensive methodology.

Figure 7B shows the operational requirements at the metric level, which offers a finer-grained view that reveals trade-offs hidden by aggregation at the evaluation criteria level. Very few metrics are universally easy to implement across all three operational requirement aspects. Metrics that have low operational requirements on all aspects were only used in one study, in which they used the method groups with the lowest operational requirements. Instead, more commonly, inverse relationships are observed between labor and the equipment aspect. For example, connectivity/fragmentation criteria have low labor scores but high equipment and cost scores. This pattern reflects the underlying methodological basis of each metric within the connectivity/fragmentation criterion: remote-sensing or modeling approaches reduce labor requirements but increase equipment and cost demands, while biodiversity metrics still rely heavily on labor-intensive ground surveys that minimize equipment needs.

This demonstrates that the choice of method predetermines the specific operational trade-offs that would drive the cost: high labor for low tech equipment, or high tech equipment for low labor. The general average cost score being positively correlated with equipment requirement shows a general trend that studies are more in favor of using more methodologies involving advanced equipment, instead of labor-intensive ones.

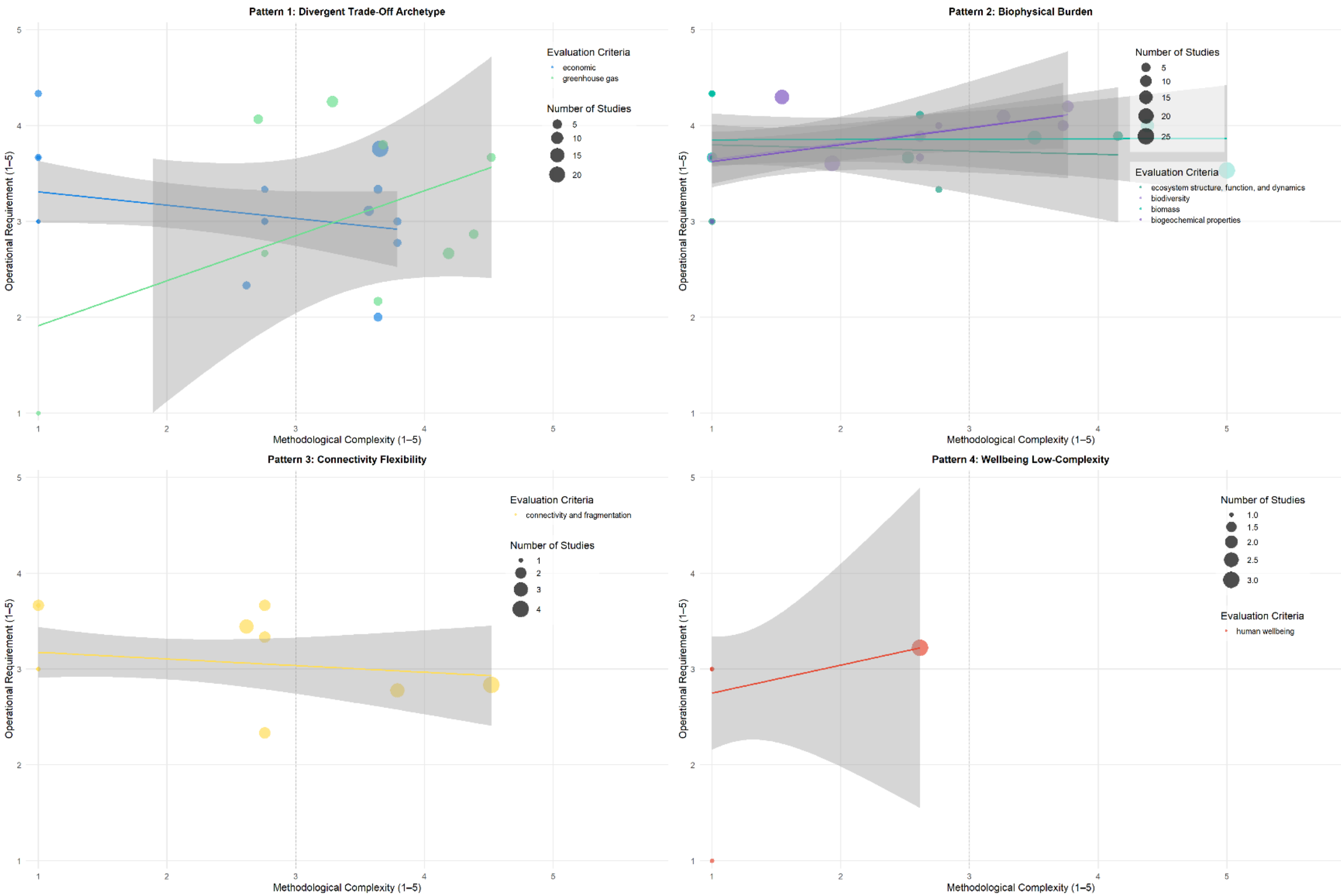

Figure 8 Patterns of methodological complexity and operational requirements relationship. The grey dashed lines divide the space into quadrants at the midpoints of each axis (score = 3). Solid lines show linear regression fits, with shaded bands representing 95% confidence intervals around the trend.

3.4. METHODOLOGICAL COMPLEXITY AND OPERATIONAL REQUIREMENTS TRADE-OFFS AND SYNERGIES

Methodological complexity and operational requirements of metrics influence their use in NCS intervention measurement, which is reflected by the systematic mapping result. By comparing these two constraints, trade-offs or synergies in balancing conceptual rigor with practical constraints when choosing a metric can be highlighted. Four unique archetypes were identified, each describing the trade-offs and synergies between their methodological complexity and operational requirements on the evaluation criteria level (Figure 8).

The first pattern shows how methodological complexity and operational requirements can be positively or negatively correlated, illustrated by greenhouse gas and economic evaluation criteria. Although both criteria are conceptually distinct, this result focuses on how the methodological approaches used in the studies shaped both evaluation criteria's methodological complexity and operational requirement pattern, as these patterns that may also extend to other concepts of a similar nature.

The second pattern describes criteria where increasing methodological complexity does not significantly reduce the operational burden, suggesting that these metrics inherently require substantial resources regardless of the chosen method. The ecosystem structure/function/dynamics, biodiversity, and biogeochemical properties all fall into this group.

The third pattern, illustrated by the connectivity/fragmentation criterion, represents a pattern where the relationship between complexity and operational requirements is nearly moderate. This suggests that for these metrics, different methods of varying complexity can be chosen without significantly impacting the overall operational burden.

The fourth pattern shows that the human wellbeing criterion forms a distinct cluster, with all metrics positioned at the lowest end of both methodological complexity and operational requirements. This reflects the limited methodological variation among wellbeing metrics have low operational requirement barriers.

4. DISCUSSION

4.1. TRENDS IN NCS MEASUREMENTS IN THE TROPICS

4.1.1. CONGO BASIN BIOME IS UNDERREPRESENTED IN THE STUDIES

Figure 3 shows that the DRC is significantly underrepresented in NCS studies, despite its importance within the Congo Basin. The low number of studies is likely due to the high cost barrier for a cost-effective NCS implementation in the DRC (Griscom *et al.*, 2020), which was exacerbated by local challenges such as political instability and poor governance (Morgan *et al.*, 2023). These conditions further restrict logistical access, weaken the development of local capacity, and hinder decentralized implementation outside the country's capital (Brown, 2017). The DRC is heavily dependent on international funding, and among the funds received, there is a lack of funds going towards NCS implementation (Brumberg *et al.*, 2025).

The minimum number of studies in the DRC means that some NCS metrics have not been tested and refined within the DRC's unique environmental and socioeconomic conditions. A metric can perform differently under different local conditions: ecological and socioeconomic-related metrics are context-dependent, meaning their performance and relationships can vary in magnitude or direction depending on local environmental conditions (Catford *et al.*, 2022 ; Edmonds, 2017). This local variability introduces layers of unknown uncertainties that limit the reliability of cross-context comparisons and reduce confidence in applying NCS outcomes from other regions directly to the DRC.

For instance, the DRC's biennial UNFCCC report lacks activity data for the AFOLU sector due to limited technical capacity (UNFCCC, 2024). A similar pattern is evident in this study, where studies measuring in situ greenhouse gases were not found (Figure 3). As a result, several emission and removal estimates were incomplete or unreported, highlighting the high uncertainty of applying standardized metrics in the DRC context. Such gaps in verifiable evidence can reduce donor confidence – once again creating cost barriers to research and climate finance (Rosenstock & Wilkes, 2021). This produces a reinforcing cycle where limited local data constrains funding, and constrained funding in turn perpetuates the lack of local data. Addressing these critical data gaps is therefore essential not only for robust metric validation but also for enabling credible and actionable NCS interventions that suit the DRC's specific environmental and socioeconomic context.

4.1.2. Co-benefit metrics as proxies to measure NCS benefits

In the NCS interventions examined, the metrics used often reflect broader goals and priorities rather than directly measuring changes in the specific ecosystem components the interventions aim to improve (as represented by the evaluation criteria in this study). Figure 4 shows that metric selection frequently does not correspond strictly to the specific ecosystem changes the intervention intends to achieve. For example, nutrient management interventions primarily target soil health improvement, yet most studies evaluate their success based on economic outcomes like crop yield rather than direct soil condition metrics. Similarly, interventions designed to avoid mangrove degradation focus ecologically on increasing biomass and reducing greenhouse gas emissions, but many assessments prioritize economic cost-benefit analyses over ecological measures.

These metrics act as proxies, which should eventually be representative in measuring the magnitude of the intervention's effect towards realizing NCS benefits. Striking a balance between measuring climate additionality and co-benefits is therefore critical, highlighting the need to assess the quality and reliability of proxies (Massi *et al.*, 2022) to ensure they accurately reflect progress toward climate, biodiversity, and human wellbeing.

4.2. Methodological complexity

4.2.1. High methodological complexity

The characteristics of metrics with high methodological complexity and a high number of studies (Figure 5) can be explained by the existence of existing standardized frameworks for each metric and clear methodological guidelines. Such guidelines offer the advantage of lowering the entry barrier for practitioners to use these metrics with the provision of an operational blueprint and reducing uncertainty in method selection. A notable example is the greenhouse gas evaluation criteria, where multiple measurement guidelines (IPCC,

2019) have led to consensus on standardized units, resulting in a limited number of total metrics (n = 9). A similar pattern is observed in biochemical soil properties, which follow established measurement protocols (Gowing *et al.*, 2020).

However, using metrics with high methodological complexity requires caution. A wide range of methods for a single metric complicates comparability, and reported values may not be directly interchangeable without understanding underlying methodologies. It is therefore important to critically assess study designs and reported uncertainty to avoid the bias of treating all values from the same metric as equivalent.

Metrics with high complexity but used in fewer studies highlight areas where methodological development is still needed. Metrics were essentially had to go through this pattern before sufficient research and institutional support made them adoptable in various studies. An example of this case is the relatively recent emergence of interest in the soil's role in climate change mitigation (Francis & Tóth, 2022 ; Jia *et al.*, 2024), where it drove the motivation of studies to measure soil metrics.

4.2.2. LOW METHODOLOGICAL COMPLEXITY

Low methodological complexity is observed for different reasons across different evaluation criteria. In the ecosystem structure/function/dynamics and biodiversity, low complexity often results from the use of many metrics that each appear in only one study and rely on a single or dominant measurement method. This reflects methodological fragmentation: many metrics capture only parts of an ecosystem, and each of these metrics is commonly measured by a unique method group. Although these metrics capture different parts of the ecosystem separately, the ecosystem itself is interconnected and interdependent (Parrott, 2010). For example, the structure of a forest (e.g., the number and types of trees) directly influences its function (e.g., carbon cycling, water filtration) and its dynamics (e.g., how it responds to a disturbance like a fire). The same case can be found in biodiversity evaluation criteria: metrics representing biodiversity ranged from genetic measurement to ecosystem diversity. Different scales of biodiversity contribute uniquely to ecosystem multifunctionality (Pasari *et al.*, 2013).

By contrast, the human wellbeing criterion shows low methodological complexity for different reasons. Having the fewest metrics (9) and studies (n = 11), human wellbeing represents a more nascent research area. In contrary to ecosystem structure/function/dynamics and biodiversity, low methodological complexity in human wellbeing evaluation criterion was caused by the limited variance of quantitative approaches in social sciences, such as surveys and questionnaires. This evidence of both limited options to measure human wellbeing, with a limited number of methods having been applied, posed a risk of failing to capture the multidimensional aspects of wellbeing, which encompass social, cultural, economic, political, health, and environmental factors, as well as considerations of equity and local context (Betley *et al.*, 2023).

4.3. OPERATIONAL REQUIREMENT

4.3.1. VARIATIONS OF METHOD PREFERENCES

Figure 6 provides the foundation for understanding the operational profiles of each evaluation criterion. It shows that specific criteria are primarily measured using a distinct set of methods, and these methods carry predictable resource burdens. The trend shows that metrics are being measured through either technologically advanced methods, or labor-intensive methods.

Method approaches in which cost is aligned with equipment requirements are laboratory analysis and remote sensing. Laboratory analysis is related to the infrastructure needed to conduct measurements that cannot be measured in situ. Access to laboratory infrastructure has a layer of cost, which not only includes the computational equipment itself but also the cost to transport the sample from the study site to the laboratory. In this study, laboratory analysis was used mostly to measure soil-related metrics. Out of the 102 studies that use laboratory analysis, including the combination that comes with it, 56 studies were used in soil-related metrics. This includes, among others, the Dumas dry combustion method (FAO, 2020) or chloroform fumigation-extraction method for soil microbial biomass (FAO, 2024). In contrast, 16 studies that measured soil-related metrics did not use laboratory access and solely used direct in situ measurements, such as the use of visual survey (Mulimbi *et al.*, 2023) or handheld measurement devices (Gonçalves Oliveira *et al.*, 2024).

As for remote sensing, advanced equipment is needed to access the raw imageries both through active and passive sensors, until the data processing process. This method is used mostly for connectivity/fragmentation (n = 7) to produce metrics such as land cover changes or forest connectivity indices. However, operational requirement increases when it is combined with other methods, in particular, the combination with ground visual survey for ground truthing process due to the increase in required personnel. The use of this method combination is common for metrics related to biomass evaluation criteria (n = 4).

On the other hand, the two most labor-intensive methods are ground visual survey and social data collection. Biodiversity evaluation criteria account for 25.5% (n = 34) among all ground visual survey measurements, followed by biomass (n = 31). Despite the rise of advanced technologies, biodiversity and biomass still rely heavily on ground measurements because these methods provide critical, high-resolution data that remote sensing cannot capture. While remote sensing is excellent for mapping broad patterns of land cover change or estimating biomass across large areas, it cannot identify species with certainty or provide fine-grained ecological details (Cavender-Bares *et al.*, 2022). Similarly, for biomass, field inventories are still conducted as they are important region-specific biomass models (De Almeida *et al.*, 2025).

The evidence shows that method preferences are often criterion-specific: soil metrics are tied to laboratory analysis, connectivity/fragmentation to remote sensing, and biodiversity and biomass to ground surveys. These distinct patterns highlight that methodological choices are not universally applicable across criteria but instead are shaped by the distinct operational and data requirements of each measurement type.

4.3.2. TRADE-OFFS IN METHOD SELECTION

Figure 7 shows the trade-off of the operational requirements of each evaluation criterion, where they are mostly either labor-intensive or technologically advanced. Both options have their own considerations. Labor-intensive metrics are often socially beneficial, particularly in regions where job creation and community empowerment are priorities. These approaches create employment opportunities at multiple skill levels, strengthen resilience, and support long-term economic recovery (Chausson *et al.*, 2024). In contrast, high technology methods are often less labor-intensive and highly scalable (Camarretta *et al.*, 2020). High technological approaches can cover vast landscapes quickly and are cost-effective. The main drawback is the need for significant upfront investment and specialized technical expertise, which can make them less feasible in low-capacity settings with limited infrastructure and funding. For example, advanced measurement methods such as modeling and remote sensing, their use in African restoration projects is often limited by inadequate internet, unreliable electricity, and the need for specialized training and funding (Elias *et al.*, 2025). In addition, the viability of high-tech methods often relies on the availability of reliable data, much of which is generated or owned by local stakeholders. If such data is used without meaningful participation, communities risk being excluded from decisions affecting their resources, whereas active involvement in managing and interpreting data can turn high-tech tools into avenues for empowerment and capacity building (Phillipson *et al.*, 2012).

In summary, the effort to lower the operational cost comes with three trade-off options: job creation, scalability, or data quality. The choice of trade-off is often context-specific, as feasibility depends on the particular conditions of each setting (Elias *et al.*, 2025). Scaling labor-intensive approaches to vast landscapes, such as large forest inventories and destructive sampling campaigns before remote sensing, would be prohibitively expensive and logistically difficult. However, studying agroforestry impact in DRC with labor-intensive methods proved effective for capturing detailed socioeconomic and biophysical data (e.g., Kachaka *et al.*, 2020). Though less scalable, these methods were crucial for ensuring projects delivered genuine community benefits and for validating ecological restoration at fine scales.

4.4. METHODOLOGICAL COMPLEXITY AND OPERATIONAL REQUIREMENT TRADE-OFFS AND SYNERGIES

Each evaluation criterion possesses a distinct pattern in the relationship between methodological complexity and operational requirements (Figure 8). Four archetypes were identified describing the trade-offs between methodological complexity and operational requirements. While each pattern is illustrated by the set of evaluation criteria defined in this study, their underlying characteristics can be applied more broadly – showing how a metric measurement could follow certain archetypes.

Pattern 1: Divergent trade-off archetype

Pattern 1 illustrates two contrasting archetypes in how methodological complexity relates to operational requirements. These divergent patterns highlight that evaluation criteria could possess a distinct trade-off structure based on the nature of their data and methodological foundations.

For greenhouse gas evaluation criteria, the positive trend reflects the common tiering system of this evaluation criterion. Simpler metrics, such as avoided emissions (complexity: 3.6, operational requirement: 2.2), can be applied with limited resources, whereas metrics like Global Warming Potential (complexity: 4.5, operational requirement: 3.7) require broader datasets and methodological integration (Cherubini *et al.*, 2016). Other metrics, such as N_2O emissions (complexity: 3.3, operational requirement: 4.25), illustrate that high methodological complexity is equated with high operational requirements due to requiring field sampling or specialized measurement techniques. This shows how higher-tier metrics are considered more complex because they rely on more detailed, granular data. In contrast, simpler metrics can provide useful estimates with lower operational effort, but moving to higher tiers requires more sophisticated methods and greater resources to capture finer detail and reduce uncertainty.

The opposite trend is shown in the economic evaluation criterion. In this case, the high complexity of the metrics was largely driven by methods that rely on computational modeling, mathematical optimization, and the use of secondary data, which kept the average operational expenditure relatively low. Examples from the study include variations of cost-benefit metric measurement (complexity: 3.6, operational requirement: 2) using greenhouse gas benefits monetization from secondary market value (Cameron *et al.*, 2019), forecasting cash flows with discounted cash flow models (Lucchesi *et al.*, 2024), and optimizing restoration sites using a Gurobi optimizer (Tan *et al.*, 2022). In these instances, analytical sophistication does not necessarily translate into high operational costs because the primary input is often existing data sources rather than generating new data. However, realizing these low-cost, high-complexity metrics requires an initial, potentially resource-demanding research and development phase, such as gathering primary crop yield data for a reference value (complexity: 3.65, operational requirement: 3.75).

Together, these examples show how different evaluation criteria – beyond greenhouse gas and economic – can embody distinct archetypal relationships between complexity and operational demands. In the case of greenhouse gas metrics, methodological rigor is tightly bound to increased operational effort, whereas in economic metrics, complexity is often decoupled from ongoing operational requirements. Recognizing this archetype helps practitioners anticipate resource needs and make informed trade-offs between rigor and feasibility.

Pattern 2: High operational burden of biophysical metrics

Ecosystem structure/function/dynamics, biogeochemical properties, biodiversity, and biomass possess a relatively high operational requirement but range from low to high methodological complexity. This group of metrics has a relatively wide range of method approaches, yet they are costly. As a result, this archetype reflects not whether operational requirements can be minimized, but rather how those operational burdens are distributed amongst the methodological approaches, whether they are technologically intensive or labor-intensive. This result highlights the complexity of the physical environment which requires investments that cannot be bypassed without compromising data quality (Moats *et al.*, 2022 ; Bellocchi *et al.*, 2010).

Pattern 3: Connectivity/fragmentation metrics' flexibility

Connectivity/fragmentation metrics are evenly distributed across the spectrum of methodological complexity, while maintaining moderate operational requirements. This group shows a pattern of where varying technological advancements of operational requirements could have a wide range of metric complexity options.

Landscape-scale monitoring, using metrics within the connectivity/fragmentation group, mostly relying on advanced sensing technologies, but labor-efficient. This reflects the benefits of digital data processing pipelines where large-scale, high-resolution datasets (e.g., satellite imagery) replace repeated, on-the-ground measurements. The range of methodological complexity, from very low (H'=0) in standardized, single-method metrics like land use changes or tree cover loss (Rother *et al.*, 2018), to high ones in composite metrics like intervention priority areas, which integrate multiple spatial layers and decision rules (Cavalcante *et al.*, 2022). This spread suggests a spectrum: some metrics benefit from mature, streamlined workflows with minimal methodological diversity, while others remain in an innovative phase where complexity is driven by integrating diverse data sources and algorithms.

The trend of methods used to measure these metrics (i.e., remote sensing and modeling, Figure 5) serves as an opportunity to act as an intermediate information to improve the scale and quality of data, while maintaining its cost-effectiveness – especially related to the problem addressed within the first two patterns. Recent advances in remote sensing highlight that combining field inventories with high-resolution imagery and scalable digital platforms can enhance regional monitoring models. In this sense, connectivity/fragmentation metrics show how moderate operational costs, when supported by cloud computing pipelines and machine learning techniques, can produce scalable solutions that improve data precision while minimizing repetitive ground-based efforts (De Almeida *et al.*, 2025).

Pattern 4: Human wellbeing is a less-explored low-complexity metric

Human wellbeing is the least used evaluation criterion to measure NCS outcomes, with minimal method diversity among those applied. The most common measure, subjective wellbeing, appears in only 3 studies, while other metrics appear just once. With so few studies, developing established methods is challenging, creating a cycle where limited research prevents methodological development.

Of the eleven studies measuring wellbeing, six focus on Indonesian agroforestry systems. This may reflect the long-known benefits of these systems, linking tree cover with livelihoods, food security, and community resilience. However, the narrow geographic and thematic scope makes it unclear whether these measurements apply more widely. Without broader evidence, human wellbeing remains a marginal part of NCS assessment, risking the neglect of whether climate solutions truly benefit the people they aim to serve.

4.5. STUDY LIMITATIONS AND FOLLOW-UP

This study's research questions depend heavily on the details of studies describing their methodology. Many studies do not fully describe the reason behind their methodological choices, underlying assumptions, or treatment of uncertainty. As a result, this study's analysis framework was designed to address this limitation, which introduces a degree of subjective interpretation. To compensate for this limitation, the workflow to generate this study's result is made fully transparent, replicable, and modular to support refinements if more information were to be provided.

This study does not involve grey literature as a part of the scoped literature. While this exclusion might introduce publication bias to the results, the decision to exclude grey literature was mainly due to the limited resources to conduct the grey literature screening process as it requires a different approach than to screen published literature (Paez, 2017). Nevertheless, grey literature, such as project reports and white papers, is a valuable source of information to be added in this context as it covers implementation evidence of various NCS activity measurements.

Another further work that can be done is to extract and examine additional variables influencing the robustness of reported outcomes. These include the temporal perspective of metrics (short-term versus long-term monitoring), the spatial scale of measurement (local, regional, or global), and contextual conditions such as governance structures, socio-economic settings, or ecosystem types. Incorporating these dimensions into future analyses would provide a more nuanced understanding of how NCS performance varies across contexts and over time.

A valuable follow-up would be to systematically analyze the quality of proxy metrics in relation to the benefits that NCS aims to deliver. Many studies rely on indirect indicators, such as canopy cover as a proxy for carbon storage or household income as a proxy for human wellbeing. Future research should test how accurately these proxies capture intended NCS benefits and identify where reliance on proxies may misinterpret outcomes or overestimate benefits.

5. CONCLUSION

This study shows that no single pattern defines how metrics are used to measure NCS interventions. Their application depends on the intervention context, shaped by methodological complexity and operational requirements. The DRC's underrepresentation highlights the context-dependent nature of metrics, where ecological and socioeconomic conditions influence performance and cross-country comparability, and gaps in local data reinforce cycles of limited funding and constrained measurement. NCS interventions often measure co-benefits instead of direct ecosystem components' change, which highlights the need for intermediate metric proxies that can quantitatively prove these co-benefits.

Methodological complexity is not uniform across evaluation criteria and metrics but reflects the scientific maturity of a field and the conceptual demands of the metric itself. High-complexity metrics, such as biomass, benefit from well-established, multi-scale frameworks and standardized protocols, enabling comparability and signaling global relevance. Low-complexity metrics often reflect fragmentation, where broad concepts are reduced to narrow methods or single-scale indicators – highlighting both methodological immaturity and limited adoption.

Operational requirements vary not only by evaluation criterion but also by trade-offs stakeholders are willing or able to make between cost, scalability, data quality, and social benefit. Technological approaches such as laboratory analysis and remote sensing provide scalability and comparability but demand infrastructure, expertise, and investment. Labor-intensive methods, by contrast, may be more beneficial in low-capacity settings and can deliver critical socio-economic and ecological insights, though scalability remains a problem.

Taken together, four patterns of metrics by their operational requirement and methodological complexity identified in this study illustrate that some metrics require unavoidable investments in labor or technology, while others are able to achieve methodological complexity with relatively low operational costs, and how human wellbeing is underdeveloped and at risk of marginalization. Importantly, connectivity/fragmentation metrics show that advances in remote sensing and digital processing pipelines can offset the high operational costs of obtaining biophysical data, enabling scalable monitoring at moderate cost – even when top-tier ecological data demands high methodological complexity and resource intensity. These findings suggest that effective design of NCS intervention measurement requires recognizing trade-offs between complexity and operational requirement, while also leveraging methodological advances to lower barriers and identifying which metrics are likely to be adopted, need targeted investment, or risk neglect despite their ecological and social importance.

.

6. REFERENCES

Bellocchi G., Rivington M., Donatelli M., & Matthews K. 2010. Validation of biophysical models: issues and methodologies. A review. *Agronomy for Sustainable Development*, 30(1), p. 109-130. DOI : 10.1051/agro/2009001

Betley E.C., Sigouin A., Pascua P., Cheng S.H., MacDonald K.I., Arengo F., Aumeeruddy-Thomas Y., Caillon S., Isaac M.E., Jupiter S.D., Mawyer A., Mejia M., Moore A.C., Renard D., Sébastien L., Gazit N., & Sterling E.J. 2023. Assessing human well-being constructs with environmental and equity aspects: A review of the landscape. *People and Nature*, 5(6), p. 1756-1773. DOI : 10.1002/pan3.10293

Brown H.C. 2017. Implementing REDD+ in a Conflict-Affected Country: A Case Study of the Democratic Republic of Congo. *Environments*, 4(3), p. 61. DOI : 10.3390/environments4030061

Brumberg H., Hegwood M., Eichhorst W., LoPresti A., Erbaugh J.T., & Kroeger T. 2025. Global analysis of constraints to natural climate solution implementation. *PNAS Nexus*, 4(6), p. pgaf173. DOI : 10.1093/pnasnexus/pgaf173

Camarretta N., Harrison P.A., Bailey T., Potts B., Lucieer A., Davidson N., & Hunt M. 2020. Monitoring forest structure to guide adaptive management of forest restoration: a review of remote sensing approaches. *New Forests*, 51(4), p. 573-596. DOI : 10.1007/s11056-019-09754-5

Cameron C., Hutley L.B., Friess D.A., & Brown B. 2019. High greenhouse gas emissions mitigation benefits from mangrove rehabilitation in Sulawesi, Indonesia. *Ecosystem Services*, 40, p. 101035. DOI : 10.1016/j.ecoser.2019.101035

Catford J.A., Wilson J.R.U., Pyšek P., Hulme P.E., & Duncan R.P. 2022. Addressing context dependence in ecology. *Trends in Ecology & Evolution*, 37(2), p. 158-170. DOI : 10.1016/j.tree.2021.09.007

Cavalcante R.B.L., Nunes S., Viademonte S., Rodrigues C.M.F., Gomes W.C., Ferreira J.D.S., Pontes P.R.M., Giannini T.C., Awade M., De S. Miranda L., & Nascimento W.R. 2022. Multicriteria approach to prioritize forest restoration areas for biodiversity conservation in the eastern Amazon. *Journal of Environmental Management*, 318, p. 115590. DOI : 10.1016/j.jenvman.2022.115590

Cavender-Bares J., Schneider F.D., Santos M.J., Armstrong A., Carnaval A., Dahlin K.M., Fatoyinbo L., Hurtt G.C., Schimel D., Townsend P.A., Ustin S.L., Wang Z., & Wilson A.M. 2022. Integrating remote sensing with ecology and evolution to advance biodiversity conservation. *Nature Ecology & Evolution*, 6(5), p. 506-519. DOI : 10.1038/s41559-022-01702-5

Chang C.H., Erbaugh J.T., Fajardo P., Lu L., Molnár I., Papp D., Robinson B.E., Austin K.G., Castro M., Cheng S.H., Cook-Patton S., Ellis P.W., Garg T., Hochard J.P., Kroeger T., McDonald R.I., Poor E.E., Smart L.S., Tilman A.R., Welker P., Wood S.A., & Masuda Y.J. 2024. Global evidence of human well-being and biodiversity impacts of natural climate solutions. *Nature Sustainability*, 8(1), p. 75-85. DOI : 10.1038/s41893-024-01454-z

Chausson A., Smith A., Reger R.Z.-Z., O'Callaghan B., Mori Clement Y., Zapata F., & Seddon N. 2024. Harnessing nature-based solutions for economic recovery: A systematic review. *PLOS Climate*, 3(10), p. e0000281. DOI : 10.1371/journal.pclm.0000281

Chausson A., Turner B., Seddon D., Chabaneix N., Girardin C.A.J., Kapos V., Key I., Roe D., Smith A., Woroniecki S., & Seddon N. 2020. Mapping the effectiveness of nature-based solutions for climate change adaptation. *Global Change Biology*, 26(11), p. 6134-6155. DOI : 10.1111/gcb.15310

Cherubini F., Fuglestvedt J., Gasser T., Reisinger A., Cavalett O., Huijbregts M.A.J., Johansson D.J.A., Jørgensen S.V., Raugei M., Schivley G., Strømman A.H., Tanaka K., & Levasseur A. 2016. Bridging the gap between impact assessment methods and climate science. *Environmental Science & Policy*, 64, p. 129-140. DOI : 10.1016/j.envsci.2016.06.019

Convertino M., Baker K.M., Vogel J.T., Lu C., Suedel B., & Linkov I. 2013. Multi-criteria decision analysis to select metrics for design and monitoring of sustainable ecosystem restorations. *Ecological Indicators*, 26, p. 76-86. DOI : 10.1016/j.ecolind.2012.10.005

De Almeida D.R.A., Vedovato L.B., Fuza M., Molin P., Cassol H., Resende A.F., Krainovic P.M., De Almeida C.T., Amaral C., Haneda L., Albuquerque R.W., Gorgens E., Romanelli J., Ferreira M., Salk C., Espinoza N., Silva C., Broadbent E., & Brancalion P.H.S. 2025. Remote sensing approaches to monitor tropical forest restoration: Current methods and future possibilities. *Journal of Applied Ecology*, 62(2), p. 188-206. DOI : 10.1111/1365-2664.14830

Edmonds B. 2017. The Room Around the Elephant: Tackling Context-Dependency in the Social Sciences. Dans : Johnson J., Nowak A., Ormerod P., Rosewell B., Zhang Y.-C. (éd.). *Non-Equilibrium Social Science and Policy*. Cham : Springer International Publishing, p. 195-208. Available at : http://link.springer.com/10.1007/978-3-319-42424-8_13 (Accessed20 août 2025).

Elias F., Djenontin I.N.S., Kamoto J.F.M., & Mansourian S. 2025. Accelerating forest landscape restoration monitoring in Africa: informing tangible actions from a practical perspective. *Restoration Ecology*, 33(3), p. e14366. DOI : 10.1111/rec.14366

Ellis P.W., Page A.M., Wood S., Fargione J., Masuda Y.J., Carrasco Denney V., Moore C., Kroeger T., Griscom B., Sanderman J., Atleo T., Cortez R., Leavitt S., & Cook-Patton S.C. 2024. The principles of natural climate solutions. *Nature Communications*, 15(1), p. 547. DOI : 10.1038/s41467-023-44425-2

FAO. 2024. *Standard operating procedure for soil microbial biomass (carbon): chloroform fumigation-extraction method*. FAO, Available at : http://www.fao.org/documents/card/en/c/cc9423en (Accessed25 août 2025).

FAO. 2020. *Standard operating procedure for soil total carbon - Dumas dry combustion method*. Available at : https://openknowledge.fao.org/handle/20.500.14283/ca7781en

Francis N. & Tóth G. 2022. Trend and network analysis to reveal disciplinary position of soil science and its sub-disciplines. *Agrokémia és Talajtan*, 71(1), p. 5-22. DOI : 10.1556/0088.2022.00103

Gonçalves Oliveira J.J., Gondim R.S., Távora Costa R.N., & Da Silva J.P. 2024. Effect of Biochar on Dwarf Green Coconut Orchard Yield and Irrigation Water Productivity.

Communications in Soil Science and Plant Analysis, 55(16), p. 2355-2366. DOI : 10.1080/00103624.2024.2359573

Gowing C., Suvannang N., & Caon L. 2020. *How to produce a soil sample for the proficiency testing of the Global Soil Laboratory Network (GLOSOLAN) – General instructions.* Rome : FAO, Available at : https://openknowledge.fao.org/handle/20.500.14283/ca8613en

Grames E.M., Stillman A.N., Tingley M.W., & Elphick C.S. 2019. An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods in Ecology and Evolution*, 10(10), p. 1645-1654. DOI : 10.1111/2041-210X.13268

Griscom B.W., Adams J., Ellis P.W., Houghton R.A., Lomax G., Miteva D.A., Schlesinger W.H., Shoch D., Siikamäki J.V., Smith P., Woodbury P., Zganjar C., Blackman A., Campari J., Conant R.T., Delgado C., Elias P., Gopalakrishna T., Hamsik M.R., Herrero M., Kiesecker J., Landis E., Laestadius L., Leavitt S.M., Minnemeyer S., Polasky S., Potapov P., Putz F.E., Sanderman J., Silvius M., Wollenberg E., & Fargione J. 2017. Natural climate solutions. *Proceedings of the National Academy of Sciences*, 114(44), p. 11645-11650. DOI : 10.1073/pnas.1710465114

Griscom B.W., Busch J., Cook-Patton S.C., Ellis P.W., Funk J., Leavitt S.M., Lomax G., Turner W.R., Chapman M., Engelmann J., Gurwick N.P., Landis E., Lawrence D., Malhi Y., Schindler Murray L., Navarrete D., Roe S., Scull S., Smith P., Streck C., Walker W.S., & Worthington T. 2020. National mitigation potential from natural climate solutions in the tropics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1794), p. 20190126. DOI : 10.1098/rstb.2019.0126

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), p. 7. DOI : 10.1186/s13750-018-0121-7

Höhl M., Ahimbisibwe V., Stanturf J.A., Elsasser P., Kleine M., & Bolte A. 2020. Forest Landscape Restoration—What Generates Failure and Success? *Forests*, 11(9), p. 938. DOI : 10.3390/f11090938

IPCC. 2019. *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*.

Jia L., Wang W., Zvomuya F., & He H. 2024. Trends in Soil Science over the Past Three Decades (1992–2022) Based on the Scientometric Analysis of 39 Soil Science Journals. *Agriculture*, 14(3), p. 445. DOI : 10.3390/agriculture14030445

Kachaka É.Y., Munson A.D., Gélinas N., & Khasa D. 2020. Adoption of an improved fallow practice using Acacia auriculiformis on the Batéké Plateau in the Democratic Republic of the Congo. *Agroforestry Systems*, 94(3), p. 1047-1058. DOI : 10.1007/s10457-019-00474-5

Key I.B., Smith A.C., Turner B., Chausson A., Girardin C.A.J., Macgillivray M., & Seddon N. 2022. Biodiversity outcomes of nature-based solutions for climate change adaptation: Characterising the evidence base. *Frontiers in Environmental Science*, 10, p. 905767. DOI : 10.3389/fenvs.2022.905767

Le Roux B. & Rouanet H. 2010. *Multiple correspondence analysis*. Thousand Oaks, Calif. : Sage Publications

Lucchesi A., Khanna M., Pereda P.C., Ussami K.A., Ruggiero P.G.C., Dornelas V.S., & Lallement T. 2024. Araguaia biodiversity corridor cost benefit analysis: Large scale restoration and sustainable agribusiness in Amazon and Cerrado. *Land Use Policy*, 141, p. 107122. DOI : 10.1016/j.landusepol.2024.107122

Maes S.L., Perring M.P., Cohen R., Akinnifesi F.K., Bargués-Tobella A., Bastin J. -F., Bauters M., Bernardino P.N., Brancalion P.H.S., Bullock J.M., Ellison D., Fayolle A., Fremout T., Gann G.D., Hishe H., Holmgren M., Ilstedt U., Mahy G., Messier C., Parr C.L., Ryan C.M., Sacande M., Sankaran M., Scheffer M.S., Suding K.N., Van Meerbeek K., Verbeeck H., Verbist B.J.P., Verheyen K., Winowiecki L.A., & Muys B. 2024. Explore before you restore: Incorporating complex systems thinking in ecosystem restoration. *Journal of Applied Ecology*, 61(5), p. 922-939. DOI : 10.1111/1365-2664.14614

Mahajan S.L., Tanner L., Ahmadia G., Becker H., DeMello N., Fidler R., Harborne A.R., Jagadish A., Mills M., Cairney P., Cheng S., Fariss B., Masuda Y.J., Pabari M., Tengö M., Wyborn C., & Glew L. 2023. Accelerating evidence-informed decision-making in conservation implementing agencies through effective monitoring, evaluation, and learning. *Biological Conservation*, 286, p. 110304. DOI : 10.1016/j.biocon.2023.110304

Massi K.G., Chaves R.B., & Tambosi L.R. 2022. Simple indicators are good proxies for ecological complexity when assessing Atlantic Forest restoration success. *Restoration Ecology*, 30(3), p. e13520. DOI : 10.1111/rec.13520

Mitsch W.J. 2012. What is ecological engineering? *Ecological Engineering*, 45, p. 5-12. DOI : 10.1016/j.ecoleng.2012.04.013

Moats L.T., Moore K.B., Cook M.R., Hall L.K., Gee K.L., & Campbell S.C. 2022. Comparing a low-cost bioacoustics-focused recording system with traditional acoustical measurement hardware. *The Journal of the Acoustical Society of America*, 151(4_Supplement), p. A247-A247. DOI : 10.1121/10.0011212

Morgan E.A., Bush G., Mandea J.Z., & Maraseni T. 2023. Community evaluation of forest and REDD+ governance quality in the Democratic Republic of the Congo. *Journal of Environmental Management*, 328, p. 116891. DOI : 10.1016/j.jenvman.2022.116891

Morgan E.A., Osborne N., & Mackey B. 2022. Evaluating planning without plans: Principles, criteria and indicators for effective forest landscape approaches. *Land Use Policy*, 115, p. 106031. DOI : 10.1016/j.landusepol.2022.106031

Mulimbi W., Brye K.R., Nalley L.L., & Birindwa D.R. 2023. Conservation agriculture assists smallholder farmers and their agroecosystem in the Democratic Republic of the Congo. *Agriculture, Ecosystems & Environment*, 355, p. 108597. DOI : 10.1016/j.agee.2023.108597

Nerfa L., Wilson S.J., Reid J.L., & Rhemtulla J.M. 2021. Practitioner views on the determinants of tropical forest restoration longevity. *Restoration Ecology*, 29(3), p. e13345. DOI : 10.1111/rec.13345

Ouzzani M., Hammady H., Fedorowicz Z., & Elmagarmid A. 2016. Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), p. 210. DOI : 10.1186/s13643-016-0384-4

Paez A. 2017. Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3), p. 233-240. DOI : 10.1111/jebm.12266

Parrott L. 2010. Measuring ecological complexity. *Ecological Indicators*, 10(6), p. 1069-1076. DOI : 10.1016/j.ecolind.2010.03.014

Pasari J.R., Levi T., Zavaleta E.S., & Tilman D. 2013. Several scales of biodiversity affect ecosystem multifunctionality. *Proceedings of the National Academy of Sciences*, 110(25), p. 10219-10222. DOI : 10.1073/pnas.1220333110

Paxton A.B., Riley T.N., Steenrod C.L., Puckett B.J., Alemu I. J.B., Paliotti S.T., Adler A.M., Exar L., McLean J.E.T., Kelley J., Zhang Y.S., Smith C.S., Gittman R.K., & Silliman B.R. 2024. Evidence on the performance of nature-based solutions interventions for coastal protection in biogenic, shallow ecosystems: a systematic map. *Environmental Evidence*, 13(1), p. 28. DOI : 10.1186/s13750-024-00350-5

Phillipson J., Lowe P., Proctor A., & Ruto E. 2012. Stakeholder engagement and knowledge exchange in environmental research. *Journal of Environmental Management*, 95(1), p. 56-65. DOI : 10.1016/j.jenvman.2011.10.005

Prach K., Durigan G., Fennessy S., Overbeck G.E., Torezan J.M., & Murphy S.D. 2019. A primer on choosing goals and indicators to evaluate ecological restoration success. *Restoration Ecology*, 27(5), p. 917-923. DOI : 10.1111/rec.13011

Pullin A.S., Frampton G., Livoreil B., & Petrokofsky G. 2022. *Guidelines and Standards for Evidence synthesis in Environmental Management, version 5.1*. Available at : www.environmentalevidence.org/information-for-authors (Accessed1 avril 2025).

Pyles M.V., Magnago L.F.S., Maia V.A., Pinho B.X., Pitta G., De Gasper A.L., Vibrans A.C., Dos Santos R.M., Van Den Berg E., & Lima R.A.F. 2022. Human impacts as the main driver of tropical forest carbon. *Science Advances*, 8(24), p. eabl7968. DOI : 10.1126/sciadv.abl7968

Rödl A. & Arlati A. 2022. A general procedure to identify indicators for evaluation and monitoring of nature-based solution projects. *Ambio*, 51(11), p. 2278-2293. DOI : 10.1007/s13280-022-01740-0

Rosenstock T.S. & Wilkes A. 2021. Reorienting emissions research to catalyse African agricultural development. *Nature Climate Change*, 11(6), p. 463-465. DOI : 10.1038/s41558-021-01055-0

Rother D.C., Vidal C.Y., Fagundes I.C., Metran Da Silva M., Gandolfi S., Rodrigues R.R., Nave A.G., Viani R.A.G., & Brancalion P.H.S. 2018. How Legal-Oriented Restoration Programs Enhance Landscape Connectivity? Insights From the Brazilian Atlantic Forest. *Tropical Conservation Science*, 11, p. 1940082918785076. DOI : 10.1177/1940082918785076

Shannon C. 1949. *The mathematical theory of communication*. The University of Illinois Press, Urbana, 117pp p.

Sowińska-Świerkosz B. & García J. 2021. A new evaluation framework for nature-based solutions (NBS) projects based on the application of performance questions and indicators approach. *Science of The Total Environment*, 787, p. 147615. DOI : 10.1016/j.scitotenv.2021.147615

Spellerberg I.F. & Fedor P.J. 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener'

Index. *Global Ecology and Biogeography*, 12(3), p. 177-179. DOI : 10.1046/j.1466-822X.2003.00015.x

Tan Z.D., Carrasco L.R., Sutikno S., & Taylor D. 2022. Peatland restoration as an affordable nature-based climate solution with fire reduction and conservation co-benefits in Indonesia. *Environmental Research Letters*, 17(6), p. 064028. DOI : 10.1088/1748-9326/ac6f6e

UNFCCC. 2024. *Technical analysis of the first biennial update report of the Democratic Republic of the Congo submitted on 30 December 2022* (FCCC/SBI/ICA/2023/TASR.1/COD). Available at : https://unfccc.int/sites/default/files/resource/tasr1_2023_COD.pdf

Westgate M. & Grames E.M. 2025. *synthesisr: Import, Assemble, and Deduplicate Bibliographic Datasets*. Available at : https://martinwestgate.com/synthesisr/

Westwood A.R., Hutchen J., Kapoor T., Klenk K., Saturno J., Antwi E.K., Egunyu F., Cortini F., Robertson M., Le Noble S., Wang J., Falconer M., & Nguyen V.M. 2023. A systematic map of knowledge exchange across the science-policy interface for forest science: How can we improve consistency and effectiveness? *Ecological Solutions and Evidence*, 4(1), p. e12214. DOI : 10.1002/2688-8319.12214

Wunder S., Duchelle A.E., Sassi C.D., Sills E.O., Simonet G., & Sunderlin W.D. 2020. REDD+ in Theory and Practice: How Lessons From Local Projects Can Inform Jurisdictional Approaches. *Frontiers in Forests and Global Change*, 3, p. 11. DOI : 10.3389/ffgc.2020.00011

7. LIST OF ABBREVIATIONS

AFOLU	: Agricultural, Forestry, and Other Land Uses
DRC	: the Democratic Republic of Congo
FTE	: Full Time Equivalent
MANOVA	: Multivariate Analysis of Variance
MCA	: Multi Correspondence Analysis
NCS	: Natural Climate Solutions
REDD+	: Reducing Emissions from Deforestation and Forest Degradation in Developing Countries
ROSES	: Reporting Standards for Systematic Evidence Syntheses
PICO	: Population, Intervention, Comparator, and Outcomes

8. TABLE OF TABLES

9. TABLE OF FIGURES

10. APPENDICES

APPENDIX 1: SEARCH STRING

Each of the Population, Intervention, and Outcome components are separated with an "AND" operator

Population	("brazil" OR "indonesia" OR "congo")
Intervention - purpose of NCS	("climate change" OR "climat* change" OR "adapt*" OR "mitigat*")
Intervention - individual pathway	(("avoid* forest conversion" OR "forest conservation" OR "avoid* deforestation") OR ("avoid* peat impact*" OR "peat* protection" OR "peat* conserv*" OR "tropical peatland*" OR "peat restoration" OR "peatland restoration" OR "peatland rehabilitation") OR ("avoid* mangrove loss" OR "mangrove conserv*" OR "avoid* coastal deforestation") OR ("wetland protection" OR "wetland conserv*" OR "wetland restoration" OR "wetland rehabilitation" OR "tropical wetland*" OR "avoid* wetland loss" OR "wetland ecosystem* conserv*") OR ("natural forest management" OR "sustainable forest management") OR ("avoided woodfuel" OR "reduced woodfuel harvesting" OR "fuelwood reduction") OR ("fire management" OR "wildfire management" OR "controlled burning") OR ("agroforest*" OR "trees in cropland*" OR "silvopasture") OR ("nutrient management" OR "nutrient cycl*" OR "soil nutrient*" OR "soil management" OR "biochar*") OR ("reforestation" OR "forest regeneration*" OR "reclamation*" OR "forest rehabilitation*" OR "afforestation" OR "native recoloni?" OR "forest restoration" OR "ecological restoration" OR "tree planting") OR ("mangrove restor*" OR "mangrove replant*"))
Outcome	("assess*" OR "analy?e*" OR "evaluat*" OR "monitor*" OR "measur*" OR "monitoring and evaluation" OR "monitoring, evaluation, and learning" OR "additionality" OR "baseline")

APPENDIX 2: DATABASE STRUCTURE: COLUMN DEFINITION

<table>
<tr><th>Column Variable</th><th>Definition</th></tr>
<tr><td>Intervention</td><td>Description of active intervention conducted in the study. Input as free text</td></tr>
<tr><td>Country</td><td>Country in which the intervention took place. Multiple countries mean the intervention were conducted within cross-country boundaries, where the country scope of this study (Indonesia/Brazil/DRC) is a part of the study.</td></tr>
<tr><td>NCS Category</td><td>Intervention category based on the relevant NCS pathways in the tropics (Griscom et al., 2020)
Manage
• Improved fire management
• Natural forest management
• Nutrient management
• Optimal grazing intensity
• Trees in agricultural lands
Protect
• Avoided forest conversion
• Avoided mangrove impact
• Avoided peat impact
Restore
• Reforestation
• Mangrove restoration
• Peat restoration</td></tr>
<tr><td>Methodology</td><td>Full description of method used in the study to measure outcomes, excluding any methods applied to design or implement the intervention itself. Input as free text</td></tr>
<tr><td>Method Group</td><td>Thematic group from the “Methodology” column. One metric can be measured with more than one method group.
1. Ground visual survey: observing and recording ecological or physical characteristics directly in the field. Requires access to the intervention site
2. Laboratory analysis: examination of collected samples (such as soil, water, plant tissue, or microbial material) under controlled laboratory conditions. Requires access to laboratory facility along with the infrastructure to relocate the sample from the intervention site to the laboratory facility
3. Modeling: the use of computational frameworks to simulate, estimate, or predict environmental changes
4. Remote sensing: acquisition of information about the Earth's surface or atmosphere without direct contact, typically using satellites, aircraft, drones, or other sensor platforms. This includes the image processing after the acquisition.
5. Secondary data analysis: interpreting data that was previously collected for other purposes, such as government surveys, published datasets, or research archives.</td></tr>
</table>

Column Variable	Definition
	6. **Social data collection:** collecting information systematically from individuals or communities to understand perceptions, behaviors, wellbeing, or socioeconomic impacts of the interventions
Metrics	Unit of measurement that was used in the study to report the study's outcome. One study could have more than one metrics. Some metrics may be compound metrics but conceptually represent the same information regarding a part of ecosystem's health (e.g., soil chemical properties typically include pH and macronutrients).
Metrics in Details	Elaborated description of the reported metric. Input as free text
Evaluation Criteria	Domains of an ecosystem of which metrics were measured: 1. **Biomass**: measures the amount of living organic matter 2. **Greenhouse gas**: measures the removals of greenhouse gases released to the atmosphere 3. **Biodiversity**: measures the species diversity within the ecosystem 4. **Economic**: measures the financial and livelihood benefits generated by the intervention, such as income, job creation, or profitability 5. **Biogeochemical properties**: metrics that characterize the inherent physical, chemical, and biological properties of the ecosystem 6. **Ecosystem structure, function, and dynamics**: measures vegetation characteristics, plant and ecosystem dynamics, and ecological processes that influence or reflect landscape resilience and function 7. **Human wellbeing**: measures intervention's impact on community health, safety, cultural values, or social equity and quality of life. 8. **Connectivity and fragmentation**: measures land cover's arrangement that can affect the movement of species and ecological processes. One metric can only be assigned to one evaluation criterion.
Operational Requirement	The operational requirement needed to produce the metrics measured for each methodology group, divided into three aspects of requirements: **Equipment**: the complexity level of the measurement tool • 1 point if the tool is straightforward and easy to use, with minimal operational steps and no need for specialized setup or interpretation. • 2 points if the tool has moderate complexity, requiring some understanding of its functions, calibration, or setup procedures. • 3 points if the tool is complex, involving multiple steps for setup, calibration, or operation, and may require referencing manuals or specialized training to use effectively. **Labor**: the intensity of labor needed to measure the metrics in full-time equivalent (FTE) • 1 point if the method requires less than 4 FTEs • 2 points if the method requires between 4 and 8 FTEs • 3 points if the method requires more than 8 FTEs

<table>
<tr><th>Column Variable</th><th>Definition</th></tr>
<tr><td></td><td> Cost: the total estimated monetary expense in monetary units required to measure one standardized unit (e.g., one hectare, one complete measurement cycle)

1 point if data acquisition or measurement requires minimal to no direct financial cost. Examples include freely and publicly accessible datasets, open-source tools, no-cost field materials, or methods that rely primarily on existing infrastructure without extra expense.
2 points if measurement involves moderate financial cost. This may cover routine expenses such as consumables (e.g., printing, basic sampling materials), travel costs for fieldwork or interviews, low-cost software or subscriptions, nominal fees for accessing data, or institutional resources bundled with standard operations. Affordable or alternative low-cost methods exist.
3 points if requires high financial investment. This includes proprietary or commercial datasets with substantial licensing fees, specialized hardware or sensors requiring costly purchase or rental, hiring external experts or consultants, advanced laboratory analyses, or any resource-intensive components with no feasible low-cost substitute

The average of these aspects is computed on two levels: metric-level and evaluation criteria-level. Calculation in the evaluation criteria-level is affected by the number of studies and methods in each evaluation criterion. The value is rescaled to 5 scale. </td></tr>
<tr><td>Method Complexity</td><td>Method complexity is defined by Shannon-index of each metric (fully described in Methodological complexity part in the main text). Higher variation of method indicates a higher complexity level of the metric. The value ranges from the scale 1 to 5.</td></tr>
</table>

Appendix 3: Method group's operational requirement

The score follows the scoring guideline in Appendix 2 (Operational Requirement's column description)

Method Group	Equip-ment	Labor	Cost	Reasoning
Ground Visual Survey	2	3	2	Tech ranges from simple tools to sophisticated sensors; needs trained crews which drives labour hours and travel; per-diem & transport make costs moderate.
Modeling	3	1	2	Demands high-performance computing and programming; not using primary field data for the model input; hardware and expert time give medium cost.
Ground Visual Survey; Laboratory Analysis	3	2	3	Combines intensive field sampling with lab instrumentation; total cash outlay and staff days both high.
Social Data Collection	1	2	3	In-person questionnaires; moderate interviewer time
Secondary Data Analysis	1	1	1	Effort depends on the access to the data; desktop computers suffice; no field effort or sampling purchases.
Remote Sensing	3	1	3	Requires satellites, drones or LiDAR, high-spec software and skilled image analysts; minimum labor for not using ground validation data; imagery or mission costs are high.
Ground Visual Survey; Modeling	3	2	3	Uses primary field data collection as the model's input
Remote Sensing; Modeling	3	1	3	Cutting-edge sensors plus model calibration; minimal labor requirement; licence and compute expenses keep cost high.
Ground Visual Survey; Social Data Collection	1	3	2	Adds interviews to in-field ecological work; normally minimum tech requirement but adds people-hours.
Laboratory Analysis	2	2	3	Bench-top instruments require specialised facilities and technicians; sampling and consumables add labour and push per-sample
Ground Visual Survey; Remote Sensing	3	2	3	Uses primary field data collection to validate the spatial analysis

Method Group	Equip-ment	Labor	Cost	Reasoning
Modeling; Social Data Collection	3	2	2	Model development plus stakeholder inputs; limited field visits; software and survey budgets are mid-range.
Ground Visual Survey; Laboratory Analysis; Modeling	3	3	3	Highest across all dimensions: fieldwork, lab instrumentation and computational modelling together demand advanced tech, large crews and substantial funds.
Remote Sensing; Social Data Collection	2	2	3	Stakeholder interviews are used as the "ground truth" for spatial data analysis
Laboratory Analysis; Social Data Collection	3	1	3	Social data collection acts as a supporting data

APPENDIX 4: LIST OF IDENTIFIED METRICS

Evaluation Criteria	Metrics	Number of studies
Biodiversity	Vegetation species diversity	22
	Soil biological properties	10
	Bird diversity	3
	Invertebrates diversity	3
	Vegetation functional diversity	3
	Macrozoobenthic diversity	2
	Bird acoustic indices	1
	Genetic diversity	1
	Seed diversity	1
	Species-interaction networks	1
Biogeochemical properties	Soil physicochemical properties	18
	Soil chemical properties	14
	Soil physical properties	9
	Overall soil quality	8
	Peat hydrology	3
	Microclimatic variables	2
	Peat soil properties	2
	Peat water table depth	2
	Forest hydrology processes	1
	Peat depth	1
	Pollutant content	1
	Streamflow	1
	Water availability	1
Biomass	Aboveground biomass	25
	Total carbon stock	16
	Soil organic carbon	14
	Basal area and density	7

Evaluation Criteria	Metrics	Number of studies
	Dry matter production	3
	Litterfall production	3
	Aboveground biomass recovery dynamics	2
	Cover crop biomass, decomposition, nutrient cycling	2
	Soil organic matter	2
	Bioenergy yield	1
	Biomass energy storage	1
	Carbon stock dynamics	1
	Components of net primary productivity (NPP)	1
	Volumetric stock	1
Connectivity and fragmentation	Intervention priority area	4
	Land cover area	3
	Vegetation cover	3
	Agroforestry spatial distribution	2
	Forest connectivity	2
	Forest fragmentation indices	2
	Land cover changes	2
	Avoided forest cover loss	1
	Biophysical suitability	1
	Coastal shoreline change	1
	Improved fallow-use area intensity	1
	Land use allocation proportion	1
	Land use changes	1
	Species distribution	1
	Tree cover loss	1
Economic	Crop yield	22
	Ecosystem service value	6

Evaluation Criteria	Metrics	Number of studies
	Cost-benefit	4
	Livestock productivity	4
	Cost effectiveness	3
	Economic benefit	3
	Economic performance	3
	Economic trade-off	2
	Forage production	2
	Household income	2
	NTFP value	2
	Timber production	2
	Cost abatement	1
	Cost of production	1
	Farmer species preference	1
	Growth model of valuable tree species	1
	Land productivity	1
	Livestock psychological parameters	1
	Management effectiveness tracking tool score	1
	Opportunity cost	1
Ecosystem structure, function, and dynamics	Plant growth and survival	11
	Vegetation structure	6
	Thermal comfort indices	3
	Deforestation data	2
	Natural regeneration potential	2
	Plant biometric variable	2
	Canopy cover	1
	Coastal exposure levels to hazards	1
	Drought resistance	1
	Fire behavior	1

Evaluation Criteria	Metrics	Number of studies
	Fire matrix analysis	1
	Fire occurrence	1
	Fire seasonality, burned area, fuel loads and fragmentation	1
	Forest understory logging impact area	1
	Global environmental impact	1
	Individual tree growth, recruitment, and natural mortality	1
	Litter mass and chemical composition	1
	Nematode abundance	1
	Post-fire plant reproduction	1
	Potential for fire reduction	1
	Vegetation composition	1
	Vegetation seasonality	1
	Wildfire occurrence	1
Greenhouse gas	Greenhouse gas emission	8
	NO_2 emission	8
	CO_2 emission	5
	CO_2 flux	5
	Greenhouse gas balance	5
	Avoided emission	4
	Global Warming Potential	4
	CH_4 emission	2
	Carbon footprint	1
Human wellbeing	Subjective wellbeing	3
	Community perception of socioeconomic impact	1
	Community social activities	1
	Community's standard of living	1
	Food security impact and economic revenue	1

Evaluation Criteria	Metrics	Number of studies
	Social feasibility	1
	Social impact assessment	1
	Socioeconomic indicators	1
	Sources of food	1

APPENDIX 5: STATISTICAL ANALYSIS RESULT

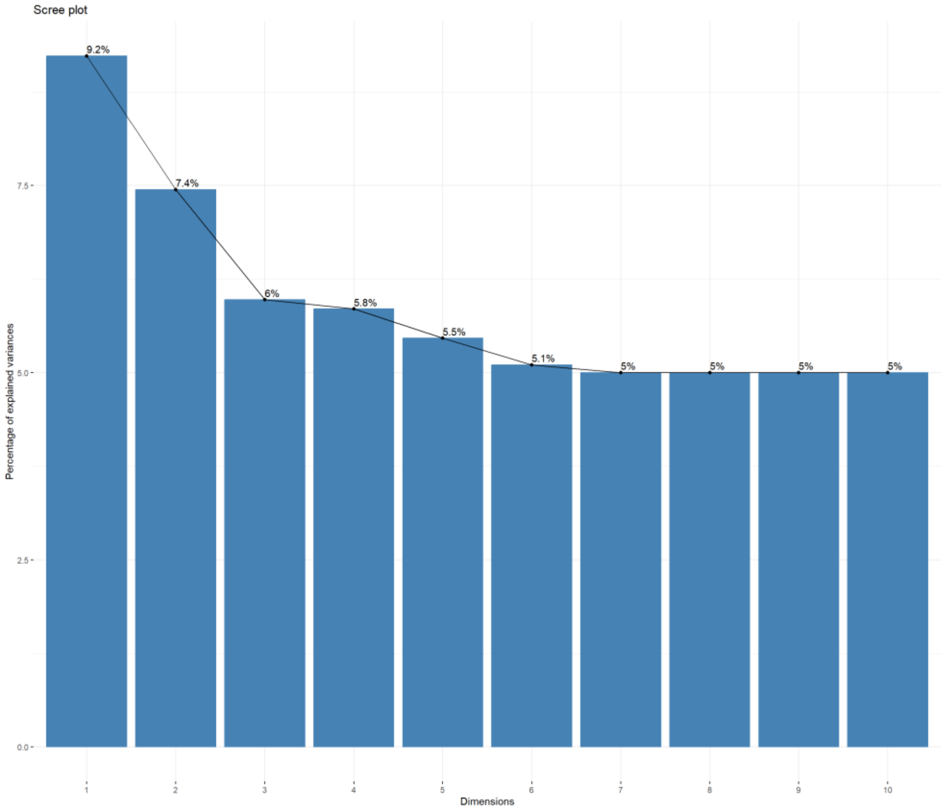

Figure 9 Scree plot of MCA. Variance is explained mostly by the first five dimensions

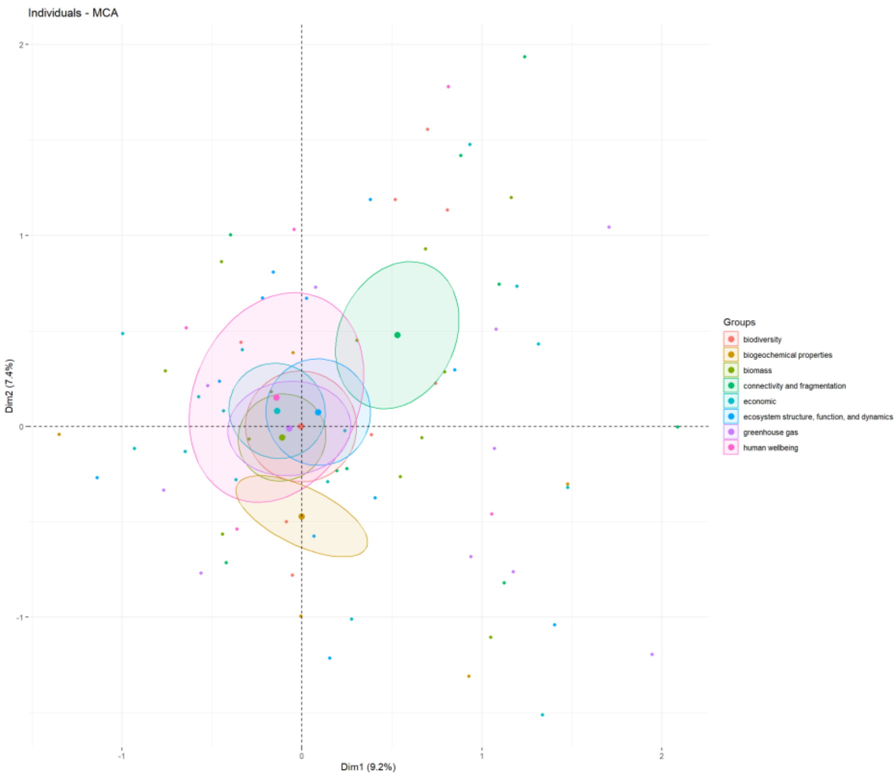

Figure 10 MCA coordinates of metrics, color coded by their evaluation criteria. The MANOVA test was applied to evaluate whether the first five dimensions of these coordinates are distinct enough based on their evaluation criteria grouping

TABLE OF APPENDICES