EUROPEAN PARTNERSHIP

How is the effectiveness of terrestrial protected areas to conserve biodiversity measured?

A systematic map

Co-funded by the European Union

Document Information

Grant Agreement number:	101052342
Project acronym:	Biodiversa+
Project full name:	The European Biodiversity Partnership
Biodiversa+ duration:	7 years
Biodiversa+ start date:	Start date: 1st October 2021
For more information about Biodiversa+	Website: http://www.biodiversa.eu/ Email: contact@biodiversa.eu Twitter: @BiodiversaPlus LinkedIn: Biodiversa+

Deliverable title:	D4.11: How is the effectiveness of terrestrial protected areas to conserve biodiversity measured? A systematic map
Authors:	Haddaway, N.R., Stoudmann, N., Alkhateeb, G., Sampson, H., Turner, K., Skidan, V., Savilaakso, S.
Work package title:	Work Package 4 Connecting R&I programs, results and experts to policy
Task or sub-task title:	Subtask 4.1.2: Desk studies and production of knowledge syntheses
Lead partner:	FRB
Date of publication:	31/03/2025
Disclaimer	Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

What is Biodiversa+

The European Biodiversity Partnership, Biodiversa+, supports excellent research on biodiversity with an impact for policy and society. Connecting science, policy and practise for transformative change, Biodiversa+ is part of the European Biodiversity Strategy for 2030 that aims to put Europe's biodiversity on a path to recovery by 2030. Co-funded by the European Commission, Biodiversa+ gathers 81 partners from research funding, programming and environmental policy actors in 40 European and associated countries to work on 5 main objectives:

1. Plan and support research and innovation on biodiversity through a shared strategy, annual joint calls for research projects and capacity building activities
2. Set up a network of harmonised schemes to improve monitoring of biodiversity and ecosystem services across Europe
3. Contribute to high-end knowledge for deploying Nature-based Solutions and valuation of biodiversity in the private sector
4. Ensure efficient science-based support for policy-making and implementation in Europe
5. Strengthen the relevance and impact of pan-European research on biodiversity in a global context.

More information at: https://www.biodiversa.eu/

Table of contents

Executive summary

Background: Protected areas (PAs) are fundamental in preserving biological diversity, supporting ecosystem services, and mitigating human impacts in today's world. However, the mere designation of PAs is insufficient for achieving conservation goals. The effectiveness of PAs needs to be ensured through employment of robust management practices and the deployment of scientifically sound monitoring methodologies.

Methods: This systematic map aims to collate and synthesize evidence from across the globe on the (monitoring) methods and metrics used to assess the effectiveness of terrestrial PAs in biodiversity conservation. By doing so, we seek to identify the consistency of monitoring schemes across different geographies and ecosystems (e.g. forested, desert, alpine and wetland biomes etc), to determine how well standardized monitoring methods have been adopted. This worldwide evidence base will not only inform the evaluation of PAs within Europe but also contribute to the global discourse on biodiversity conservation, facilitating the exchange of knowledge and best practices.

Results: The systematic map identified 275 articles assessing protected area effectiveness for primarily terrestrial PAs (i.e. a greater surface area than 50% must be terrestrial) across the globe. Most studies were conducted in Africa, Asia, and Europe, with fewer in Oceania and North America. Studies were predominantly conducted at local or regional scales, with fewer multinational assessments. Control-Intervention (CI) designs were the most common, often incorporating time-series analysis, while Before-After Control-Impact (BACI) designs remained rare. Geographic Information Systems (GIS) and direct observation were widely used, with an increasing reliance on satellite-based measurements. Studies focused largely on forested biomes, with fewer examining alpine, desert, and wetland ecosystems. Species-level assessments were limited, with many studies relying on land cover change and forest cover as a proxy for biodiversity. Genetic diversity, species traits, and ecosystem functional metrics were notably underrepresented.

Conclusions: There has been a marked increase in research on PA effectiveness in recent years. However, methodological gaps persist, particularly in the application of BACI designs and the evaluation of biodiversity beyond habitat-level metrics. The reliance on GIS and remote sensing, while useful for large-scale monitoring, may overlook species- and community-level dynamics at the local-scale. These monitoring methods may miss small-scale changes in species composition, abundance and/or richness, highlighting the complementarity of direct observation with field-based sampling methods for conservation decision making. The predominance of studies focused on a single PA reflects a tendency towards smaller-scale, site-specific assessments. While these provide detailed insights into specific locations, they may miss opportunities to evaluate broader, landscape-level impacts of PAs which could help identify connectivity issues, cumulative impacts and broader trends. Additionally, the underrepresentation of certain ecosystems and PA categories highlights the need for broader, standardized approaches to biodiversity monitoring. Addressing gaps in underrepresented biomes, such as alpine regions and deserts, could enhance our understanding of PA effectiveness across diverse ecological contexts. Future research should emphasize more diverse and integrative biodiversity indicators, improve representation across underexplored regions and ecosystems, and enhance methodological rigor in assessing PA effectiveness.

Keywords:
Biodiversity conservation, monitoring and evaluation, impact evaluation, protected area effectiveness, evidence synthesis, evidence mapping

1. Introduction

In the context of global biodiversity conservation, protected areas (PAs) are considered foundational to preserving ecological diversity, supporting ecosystem services, and mitigating human impacts (Watson et al. 2014). The strategic establishment and management of PAs are guided by international and regional policy frameworks, such as the Kunming-Montreal Global Biodiversity Framework (CBD 2022) and the European Union's biodiversity strategy for 2030 (EUR-LEX, European Union Law 2020). These frameworks set quantitative goals for PA coverage and underscore the necessity for effective and equitable management practices that are ecologically representative and well-connected (Adams et al. 2023).

However, the mere designation of PAs is insufficient for achieving conservation goals. The efficacy of these areas in biodiversity conservation is contingent upon robust management practices, the deployment of scientifically sound monitoring methodologies, and the adaptation of conservation strategies to address emergent issues (Mascia et al. 2014). The challenge of evaluating PA effectiveness toward biodiversity conservation lies in measuring direct impacts amidst complex ecological dynamics, the variability of conservation goals, and the scarcity of baseline data and long-term monitoring efforts (Geldmann et al. 2019). This endeavour is further complicated by the wide range of indicators and metrics used to gauge effectiveness, pointing to a critical need for systematic consolidation of this information (Rodrigues & Cazalis 2020).

In the context of this study, we define effectiveness as the degree to which PAs contribute towards meeting global biodiversity targets, encompassing a wide spectrum of goals, from averting species extinction to preserving threatened ecosystems (Maxwell et al. 2020). However, while the focus of this research is on biodiversity outcomes, PAs can also play a role in addressing social dimensions, such as supporting local communities, safeguarding indigenous rights, and contributing to ecological development (Borrini-Feyerabend et al. 2013). However, PAs do not inherently preserve cultural values and may have unintended outcomes. Rather, actions like supporting local communities and safeguarding Indigenous rights shape these values (Dawson et al. 2021). Nonetheless, while critically important, these social aspects fall outside the primary scope of this study.

This systematic map aims to collate and synthesize global evidence on the methods and metrics used to assess the effectiveness of terrestrial PAs in biodiversity conservation. By doing so, we seek to identify the consistency of monitoring schemes across different geographies and determine how well standardized monitoring methods have been adopted. This broad evidence base will not only inform the management and evaluation of PAs within Europe but also contribute to the global discourse on biodiversity conservation, facilitating the exchange of knowledge and best practices.

2. Objectives

The objective of this evidence synthesis was to collate and describe the evidence base relating to terrestrial protected areas and biodiversity measurement. The systematic map's primary question was:

How is the effectiveness of terrestrial protected areas to conserve biodiversity measured?

Secondary questions include:

1. Which biodiversity monitoring methods and metrics have been well used to assess protected area effectiveness, and which are under-represented or absent from the evidence base or particular contexts?
2. What is the level of consistency in monitoring schemes across geographical areas, protected area designations, and managing and researching institutions?
3. What standardised monitoring methods have been developed in the literature and how well have they been adopted?

This question can be broken down into the following key elements:

Population: Terrestrial systems globally

Intervention: Protected area establishment/presence

Comparator: Outside protected areas, before establishment of protection, or in the absence of a protected area

Outcome: Methods for measuring terrestrial protected area conservation effectiveness using direct or indirect biodiversity metrics

3. Methods

This review has been conducted according to the methods outlined in a published protocol (Haddaway et al. 2024) and based on the methods for systematic mapping published by the Collaboration for Environmental Evidence (James et al. 2016).

3.1. Deviations from the protocol

We made use of a subsampling approach to enable representative assessment of the evidence base from a manageable volume of abstracts. We manually screened 7,229 abstracts (17% of all search results) before training a machine learning model that was then used to predict the relevance of the remaining records. We then randomly sampled the results of this model, screening these additional abstracts at full text level. This resulted in a total of 655 records (19%) of the evidence base being considered at the full text screening stage. Because of our approach to sampling, we do not consider there to be any issues with a lack of representativeness or bias in the subset presented herein.

We planned to use Publish or Perish to import Google Scholar results. However, the software was not working properly at the time of use, so records were screened in situ, retaining only those that were deemed to be relevant.

3.2. Search strategy overview

We searched for evidence across a range of sources, including bibliographic databases, a web-based academic search engine, organisational websites, and using citation chasing. This range of sources aims to cover a diversity of terminological descriptions of the topic, publication platforms, traditional academic and grey literature, and citation networks of related works that might otherwise evade other search methods. The search string used in bibliographic databases was based on several previously published systematic reviews on related topics and the team's expertise in this field. The string has been tested for functionality and compared against a benchmark list (cf. Annex 1) to ensure a relevant set of primary studies are returned within the search results.

3.2.1. Bibliographic database search string

The search string adapted to Scopus was as follows:

(("protected area" OR "national park*" OR "conservation area*" OR "wilderness area*" OR "natural monument*" OR "natural park*" OR "natural feature*" OR "protected landscape*" OR "nature park*" OR "nature reserv*" OR "biosphere reserv*" OR "world heritage site*" OR "natura 2000" OR "ramsar")*
W/15
(effect OR affect* OR impact* OR effic* OR mitig* OR perform* OR success* OR indicator* OR evaluat* OR assess* OR fail* OR analy*))*
AND
(biodiversity OR ecosystem OR species OR habitat* OR communit* OR "biological diversity" OR conservation)*

3.2.2. Bibliographic databases

We searched across a suite of different bibliographic databases using the tried-and-tested search string above. The search string was adapted to the syntax of each individual resource and these search strings and the number of hits is reported in Additional File 1.

Table 1. Bibliographic databases that will be searched for relevant literature.

Database	Search details	Access/ subscription notes
Scopus	Title, abstracts, keywords	University of Tasmania
Web of Science Core Collection • Science Citation Index Expanded (SCI-EXPANDED) - 1945-present • Social Sciences Citation Index (SSCI) - 1956-present • Arts & Humanities Citation Index (AHCI) - 1975-present • Conference Proceedings Citation Index – Science (CPCI-S) - 1990-present • Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH)- 1990-present • Book Citation Index – Science (BKCI-S) - 2005-present • Book Citation Index – Social Sciences & Humanities (BKCI-SSH) - 2005-present • Emerging Sources Citation Index (ESCI) - 2015-present	Topic	University of Helsinki
CAB Abstracts through Ovid	Title OR Abstract	University of Helsinki
ProQuest (Dissertations and Theses)	Topic	University of Helsinki

3.2.3. Benchmark testing

During development of the search string, we assembled a set of 9 articles of known relevance (Annex 1) that we used as a benchmark list to test the functioning of the search string in Scopus. During testing, minor modifications were necessary to the search to retrieve all benchmark records, reflected in the final string provided above. We therefore consider this string sufficiently comprehensive.

3.2.4. Other search methods

Call for grey literature

We distributed a call for relevant grey literature via email (Biodiversa+ partners, Conservation International, IUCN, WWF, ZSL, BirdLife International) and the mailing lists of existing networks (the Collaboration for Environmental Evidence, Mongabay, the Society for Conservation Biology).

Organisational websites

We searched a range of websites of relevant organisations for relevant grey literature (see Table 2). We manually scanned the websites for publications pages and relevant literature before using built-in search facilities where available. We used the following basic search terms across all websites:

“biodiversity, protected area, national park, nature reserve”

Searches were adapted to each website depending on their successful retrieval (i.e. decreasing specificity where no results were obtained). Since search results cannot readily be exported, records are not in RIS format, and records typically lack abstracts, we screened the search results for relevance in situ (i.e. during searching), retaining only relevant records to be screened at full text along with the results of abstract screening from other sources. We recorded the search terms, number of search results, and relevant records from all website searching (see Additional File 1).

Table 2. Organisational websites that were searched for grey literature.

Organisation	URL
The World Bank Publications	https://www.worldbank.org/en/research
Conservation International	https://www.conservation.org/home
The UN Environment Programme World Conservation Monitoring Centre	https://resources.unep-wcmc.org/
International Union for Conservation of Nature	https://www.iucn.org/resources
Global Environment Facility Publications	https://www.thegef.org/newsroom/publications
International Institute for Environmental Development	https://www.iied.org/
European Commission's Joint Research Centre	https://publications.jrc.ec.europa.eu/repository/

Web-based search engine

We searched Google Scholar for additional records, which has been shown to be useful as an additional resource in obtaining grey literature as well as unindexed academic research (Haddaway et al. 2015). Because of the limitation in search functionality of Google Scholar, we used the following basic Boolean search across a variety of relevant languages restricting the searches to ‘all words in the titles’ of records on Google Scholar:

English: Biodiversity effectiveness "protected area" OR "national park" OR "nature reserve" -marine
French: Biodiversité efficacité “aire protégée” OR “parc national” OR “réserve naturelle” -marine
Spanish: Biodiversidad efectividad "área protegida" OR "parque nacional" OR "reserva natural" -marina

Portuguese: Biodiversidade eficácia, “área protegida” OR “parque nacional” OR “reserva natural” -marinha
German: Biodiversität Wirksamkeit "Schutzgebiet" OR "Nationalpark" OR "Naturreservat" -marine"
Finnish: “Luonnon monimuotoisuus” vaikuttavuus luonnonsuojelualue OR kansallispuisto OR luonnonpuisto OR erämaa-alue
Swedish: Biologisk mångfald effektivitet "skyddat område" OR nationalparker OR "naturreservat"

Although we planned to use Publish or Perish, the software was not working properly at the time of use, so records were screened in situ, retaining only those that were deemed to be relevant.

Citation chasing
We undertook forwards and backwards citation chasing, using a starting set of records of known relevance (the benchmark list of 9 articles described above) to obtain lists of research cited by and citing these records. We then deduplicated this list relative to bibliographic search results to remove records already screened, screening the remaining records for relevance. We used the tool, citationchaser (Haddaway et al. 2022), to perform citation chasing, which uses the lens.org database.

3.3. Article screening and study eligibility criteria

We screened potentially relevant records from all sources using a set of a priori inclusion criteria, outlined below. We screened records at two stages: title and abstract (together), and then full text.

We used a subsampling approach by manually screening c. 20% of the search results at title and abstract level before training a machine learning model at title and abstract level that predicted relevance of the remaining unscreened records. We then screened the full texts of all manually included records, adding a further subset of records recommended for full text screening from the machine learning model. In total, we screened c. 7% of the included abstracts at full text. In this way, we were able to obtain a representative but detailed assessment of the evidence base within the resources available for such a large topic. This machine learning and sampling approach is described in Figure 1.

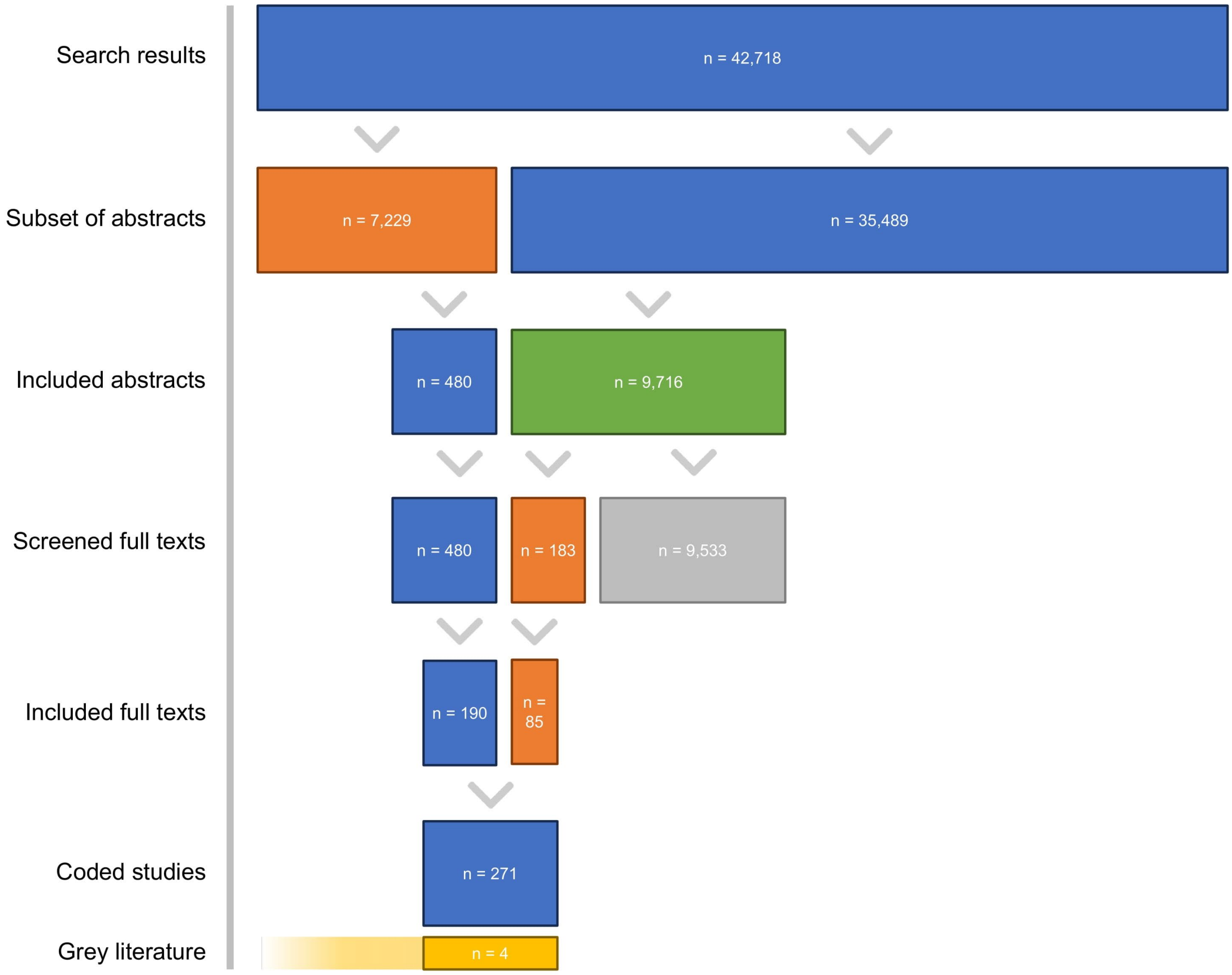

Figure 1. Graphic showing the machine learning and subsampling approach used to ensure a representative set of records is described in the map. Blue boxes represent traditional, manual approaches to screening and coding studies. The green box represents records screened using machine learning. Orange boxes represent subsamples. The grey box represents as yet unscreened but potentially relevant records.

Before commencing title and abstract screening in full, a subset of 290 records were screened independently by 4 members of the review team. Screening decisions were compared and after the first round there were 104 conflicts across all team members (67% agreement). These conflicts were discussed in depth and all conflicts were resolved through immediate discussion, with the inclusion criteria clarified where necessary. A second set of 285 records were then screened in parallel, resulting in only 10 conflicts across all 4 team members. These conflicts were again resolved through discussion, but the agreement rate of 96% was deemed to be particularly high and screening proceeded. From here on, the remaining records were shared across the team and each record was screened by one reviewer.

Full texts were retrieved using the subscription access of the combined team. Records that could not be identified or retrieved at full text were recorded and reported in Additional File 2.

All screening activities were in the review management software Rayyan.ai. None of Rayyan's AI features were used in screening.

Eligibility criteria:
The inclusion criteria for study eligibility screening are described below.

Population: Any documented research study conducted in a primarily terrestrial biome (i.e. a greater surface area than 50% must be terrestrial)
Intervention: Any officially designated protected area
Comparator: A comparison with the absence of a protected area, either prior to protection implementation, outside of the established protected area, or in a matched area lacking protection
Outcome: Any method used to measure protected area conservation effectiveness using indirect or direct quantification of terrestrial biodiversity (e.g., deforestation as a proxy indicator for biodiversity loss)
Study type: Any empirical study including primary data collection

The review focused specifically on biodiversity outcomes, excluding studies that examined ecosystem services, socio-economic impacts, cultural values, or other non-biodiversity-related outcomes. Studies that primarily investigated protected area management practices were not considered; instead, the review focused on the effects of protected area presence on biodiversity.

3.4. Data coding and presentation of results

The final set of included studies were coded, and meta-data extraction was performed manually. Meta-data included descriptive information about the location, protected area, local context, and research methods used. The data schema is described in Table 3.

Prior to commencing data extraction, the reviewers involved independently performed a pilot extraction on a set of 5 studies. Overall, data were extracted identically, but there were a small number of discrepancies in 3 out of 5 cases for essential biodiversity values that required clarification. All discrepancies were discussed in detail and the item descriptions were clarified to facilitate consistent implementation of the data schema. Once we were satisfied that the level of agreement was very high between reviewers, the remaining full texts were shared between reviewers and extracted, with each full text extracted by one reviewer. Following data extraction, the data categories were checked by an additional reviewer. All EBVs (Essential Biodiversity Variables) and survey and sampling methods were checked by at least two additional reviewers, since these proved challenging to code consistently by a single reviewer.

Box 1: PAME evaluations

Protected Areas Management Effectiveness (PAME) evaluations, which are not covered by this study due to the eligibility criteria selected, have been defined by Hockings et al. (2006) as "*the assessment of how well protected areas are being managed – primarily the extent to which management is protecting values and achieving goals and objectives*". Numerous methodologies for the assessment of PAME have been developed since the 1990s. A large part of the approaches are conceptually related to the framework from the IUCN World Commission on Protected Areas (WCPA). The Global Database for Protected Area Management Effectiveness (GD-PAME) gathers assessment data on the effectiveness of management efforts from across the globe (Coad et al., 2015) (IPBES, n.d.).

LIFE PAME Europe project

The **LIFE PAME Europe project** (2024-2028), led by EUROPARC Federation, aims to develop a methodology to be used across Europe to assess and monitor the management effectiveness (ME) in all European protected areas (PA). The interdisciplinary project consortium will take the needs of the managing authorities, the member states, and other stakeholders in the ME community into account. It aims to build on already existing systems and practices rather than starting a new one (European Commission, n.d.)

Table 3. Data schema for meta-data extraction and coding

Item	Description
Study country	Standardised country name
Study location	Free-text protected area name/description
Latitude	Provided or estimated study latitude
Longitude	Provided or estimated study longitude
Study biome	Categorical list of biome(s) from the Global Ecosystem Typology (https://global-ecosystems.org/explore)
Study ecosystem	Free-text description of the ecosystem as described by authors
Protected area name	Provided name(s) of the protected area(s)
Number of protected areas	Number of independent protected areas studied
Protection type	Provided protected area designation
Starting study year	Year study began
Study end year	Year study ended
Study duration	Sum period of time (in months) over which data were collected
Study objectives	Free-text description of the study objectives
Subject studied	Provided description of the group or taxon/taxa measured
Subject code	Categorical a posteriori (applied post hoc) groupings of subjects
Kingdom studied	Provided organism kingdom that was studied in the protected area
Sampling methodology	Categorical list of possible sampling methods: Geographical information system, Mark-recapture, Point counts, Random sampling, Sampling plots, Stratified sampling, Systematic sampling, Transects [see Annex 2 for definitions]
Survey method (outcome measurement)	Categorical list of survey method (outcome measurement) used: Active fishing, Active nocturnal search, Basket traps, Call-up survey, Camera traps, Direct listening, Direct observation, DNA extraction, Hand searching, Indirect observation, Kick sampling, Live trapping, Net fishing, Pit-traps, Quadrat sampler, Questionnaire, Radio or GPS collar, Satellite measurement, Spring-loaded bar traps, Sweep netting [see Annex 2 for definitions]
Outcome measure	Provided description of the outcome(s) measured in the study
Study type	Experimental/observational [see Annex 2 for definitions]
Experimental design	Categorical list of experimental design: BACI, BA, CI, Time-series, CI + time-series, BACI + time-series [see Annex 2 for definitions]
Scale	Categorical list of geographical scale: Local, Regional/subnational, National, Multinational
Essential Biodiversity Variables	Categorical list of Essential Biodiversity Variables studied according to https://geobon.org/ebvs/what-are-ebvs/
Comparator	Free-text description of study comparator
Funding source	Funding statements

We have summarised the evidence base in a narrative synthesis that describes patterns in the studies across settings, protected area categorisations, countries, biodiversity measurement methods, and the studied subject (e.g., species/taxa). We have used a combination of visualisations and tables of varying complexity. Heat maps present the number of studies in the evidence base categorised over 2 to 3 categorical variables. We also present the evidence base cartographically, producing an interactive visualisation displaying each study location, with descriptive information about each study in pop-ups (also known as an ‘evidence atlas’, Haddaway et al. 2018).

Using these tables and visualisations, we have sought to identify knowledge clusters (topics where sufficient evidence exists to allow meaningful meta-analysis), but primarily also knowledge gaps (areas that are un-/ or under- represented with fewer studies than might be needed/hoped/expected). By visualising methods used over various other variables (including time, geography, protected area type, taxa, etc.) we are able to identify combinations for which there are relatively fewer studies than expected. Heat maps facilitate this process by displaying the number of studies as colour graded tiles, with lighter shades indicating potential knowledge gaps and darker shades suggestion clusters. In all cases, the suitability of these topics is discussed relative to their importance conceptually as well as numerically (i.e. are they meaningful gaps).

4. Results

4.1. The review process

The flow of records through the review is outlined in a ROSES flow diagram in Figure 2. We identified over 66,000 search results from across bibliographic databases, along with 1,759 records via citation chasing from our benchmark article list. After deduplication, we were left with 42,718 unique records.

After manual and machine learning assisted title and abstract screening, we were left with 10,196 potentially relevant records to screen at full text. After subsampling (combining the 480 included abstracts from manual title and abstract screening and 183 randomly sampled records from machine learning-screened titles and abstracts), we obtained 663 full texts, which were screened manually. We identified 4 relevant full texts through grey literature searching, resulting in a total of 275 articles and 280 studies included in the final systematic map database. See Additional File 3 for the database of included studies.

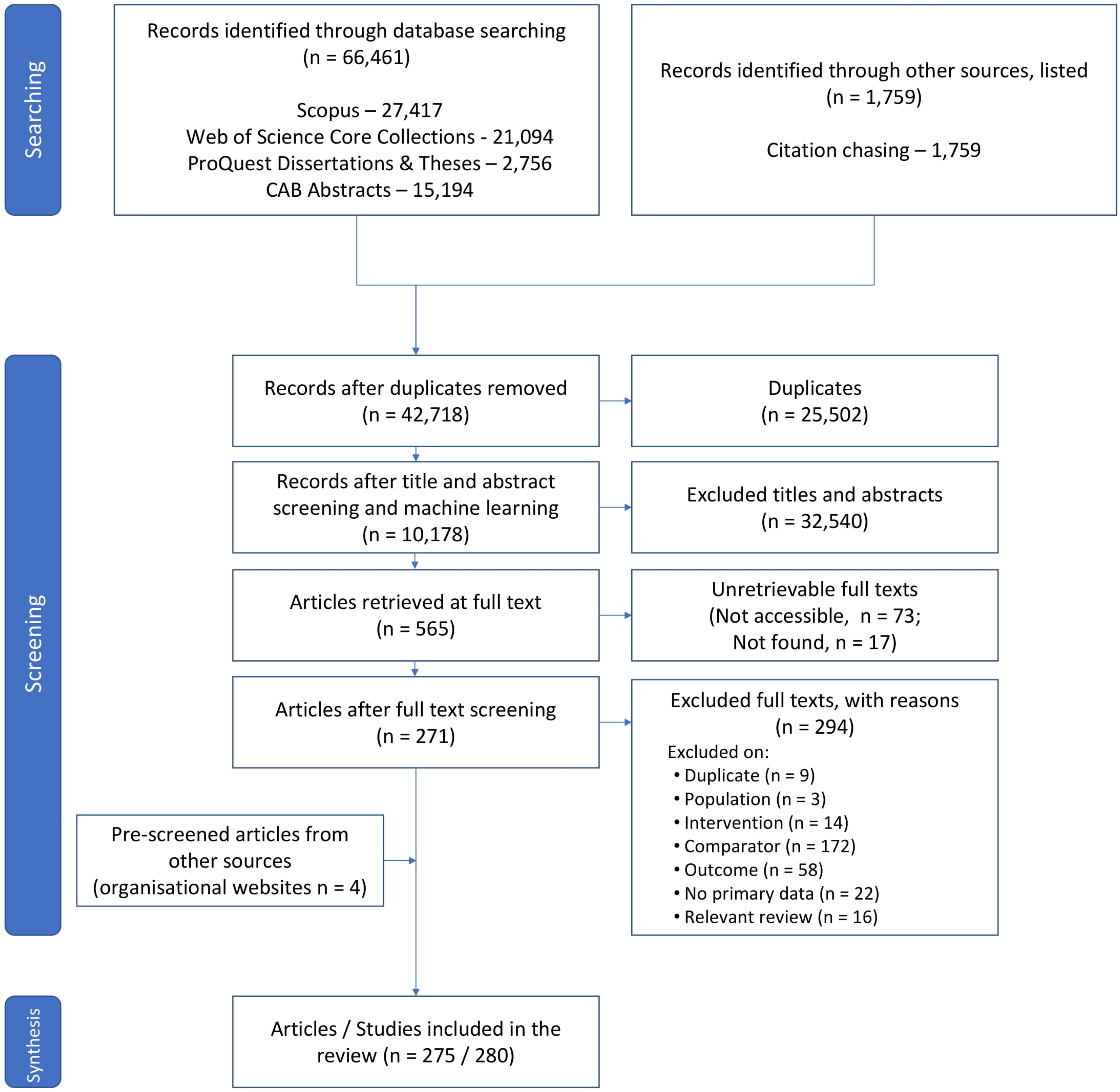

Figure 2. Adapted ROSES flow diagram showing the fate of all records included in the systematic map.

The numbers presented in the following figures are based on a subsample of the evidence base (c. 19% of all relevant full texts).

4.2. Geographical distribution of the evidence base

We have summarised the evidence base in an interactive evidence atlas (see the snapshot in Figures 3 and 4). The studies covered a range of scales. Most were conducted at the local (105) or regional/subnational scale (105). National-scale studies accounted for 39 studies, while 36 studies were multinational.

How is the effectiveness of terrestrial protected areas to conserve biodiversity measured?

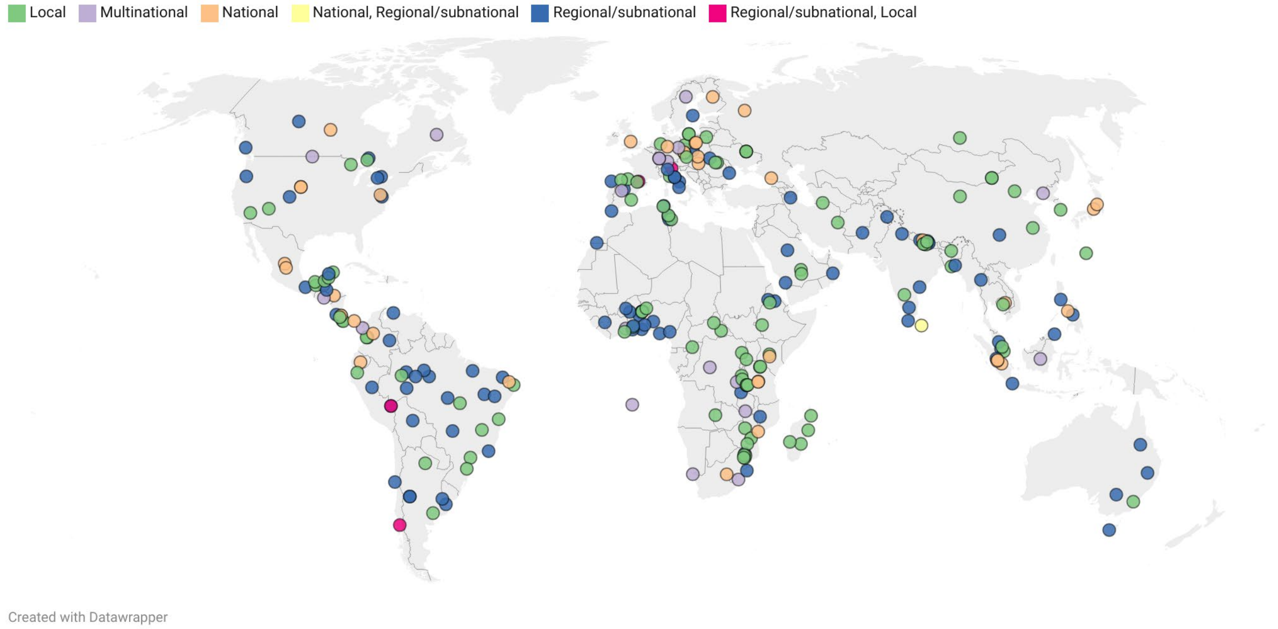

Figure 3. Evidence atlas showing the geographical location of all studies included in the systematic map, including the study scale (interactive version available here: https://datawrapper.dwcdn.net/Th2Jd/1/). Studies without latitude or longitude are displayed in the central Atlantic Ocean. Fifteen regional studies without coordinates are not shown.

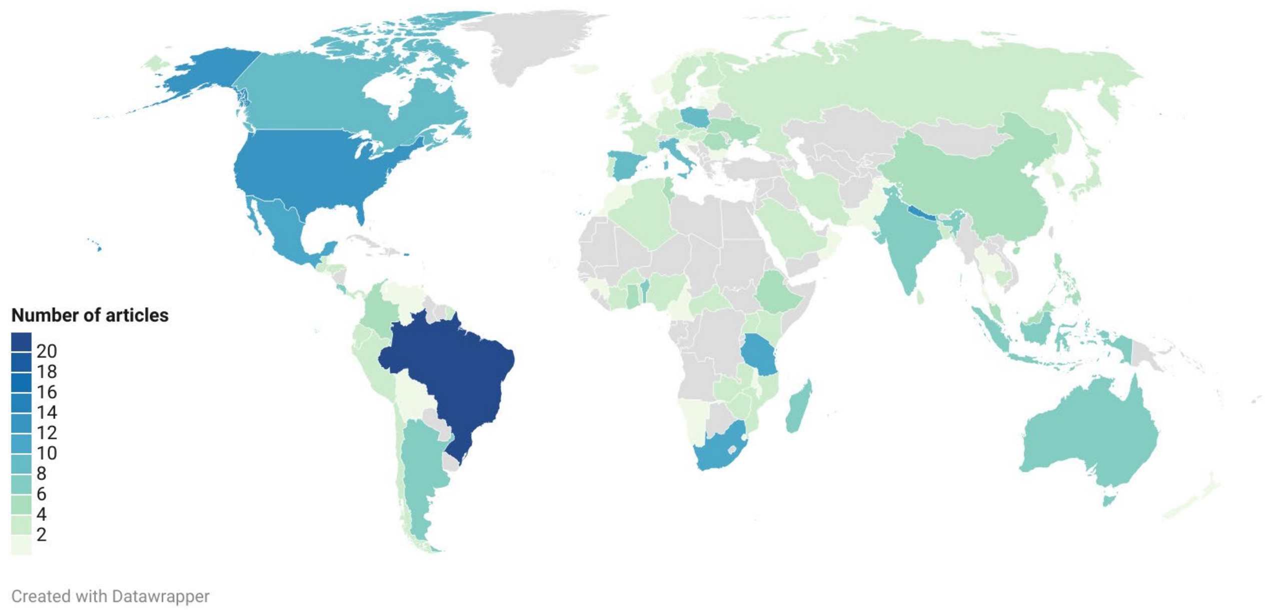

Figure 4. Choropleth showing the frequency of studies represented within the systematic map across countries. Not shown are the studies covering regions as a whole - Africa, 7: Antarctica, 1: Asia, 4: Europe, 7: Global, 8: North America, 1: Oceania, 1: South America, 4.

The geographical distribution of articles was uneven, with the largest share conducted in Africa (63 articles), followed by Asia (50 articles) and Europe (46 articles) (Table 4). South America was represented by 40 articles, while Central America (21 articles) and North America (19 articles) had moderate representation. Oceania (6 articles), MENA (Middle East and North Africa) (14 articles), and Antarctica (1 article) had limited coverage. A small number of articles (17 articles) spanned multiple regions.

Table 4. Articles per continent. SSA = Sub-Saharan Africa. MENA = Middle East and North Africa.

Continent	Articles
SSA	63
Asia	50
Europe	46
South America	40
Central America	21
North America	19
Multiple	17
MENA	14
Oceania	6
Antarctica	1

4.3. Publication trends and study duration characteristics

Very few articles published pre-2000 were identified (Figure 5). Between 2000 and 2011, there was a slight increase, particularly after 2006, through the growth remained modest. From 2012 onwards, however, the number of articles published rose sharply, with a nearly continuous upward trend.

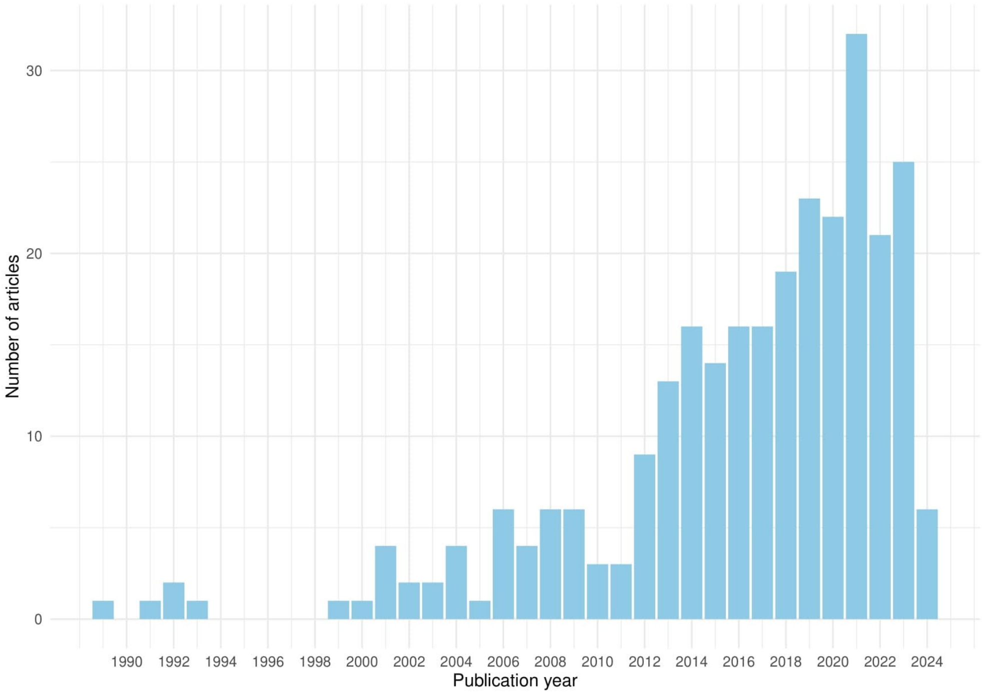

Figure 5: Number of articles published per year.

Most studies were short-term, with the majority lasting less than one year (Figure 6). A significant proportion spanned durations of up to 20 years. Studies extending up to 45 years were much less common, and only a very small number exceeded 50 years.

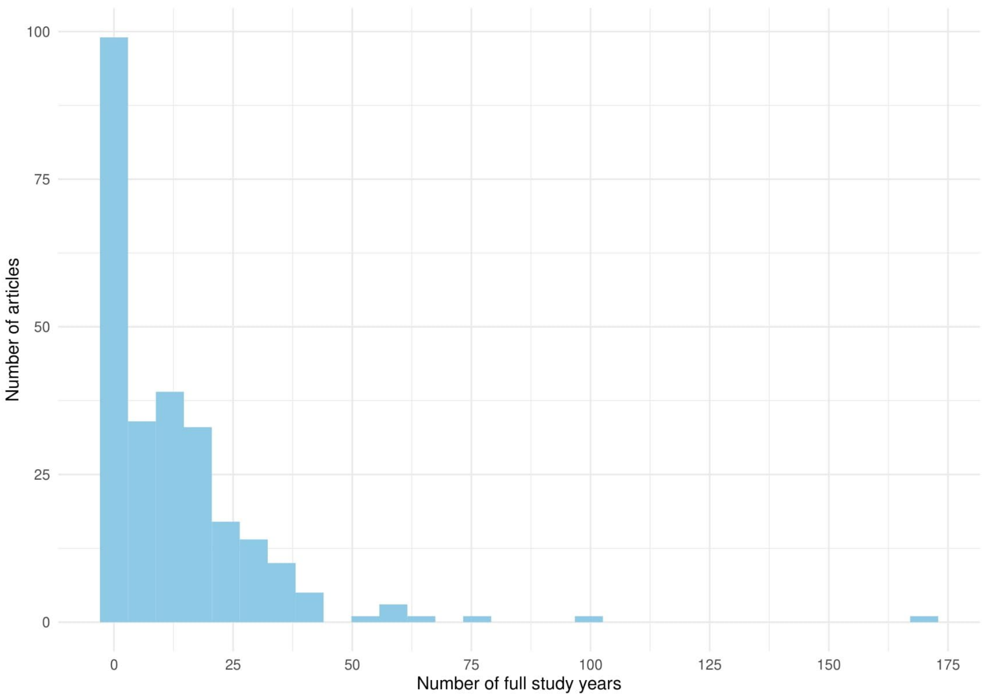

Figure 6. Number of full study years across articles (12 months or more between the year a study commenced and the year it was completed, 0 = less than one year).

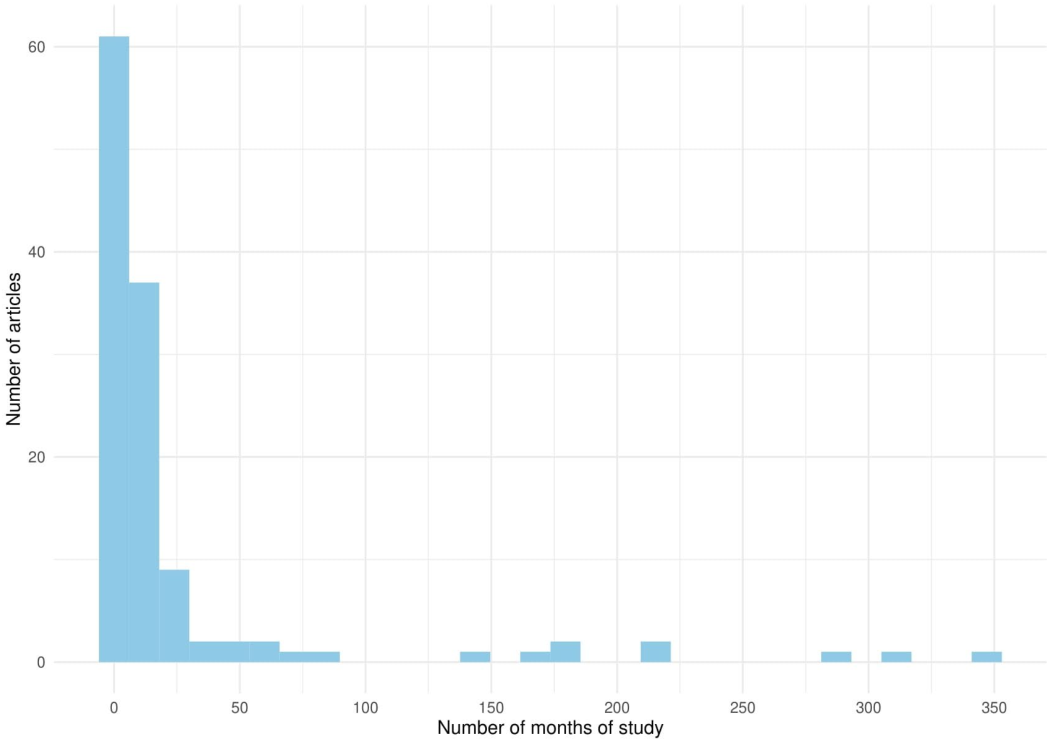

Figure 7. Number of study months across articles (total number of months in which data were recorded).

When considering the number of months of study, most lasted less than one month, and the second largest share lasted under a year (Figure 7). Beyond this, the numbers decline sharply, with only a few long-term studies extending beyond 12 months.

4.4. Study attributes and methodological approaches

Most studies focused on a single PA (Figure 8). Among studies that examined multiple PAs, a small but notable subset focused on three PAs, with numbers gradually decreasing as the number of PAs per study increased. Despite this decline, studies encompassing up to 150 PAs were represented. Beyond this range, the numbers diminished significantly, with only a handful of studies exceeding 150 PAs. The largest study included in the evidence base considered 40,232 PAs.

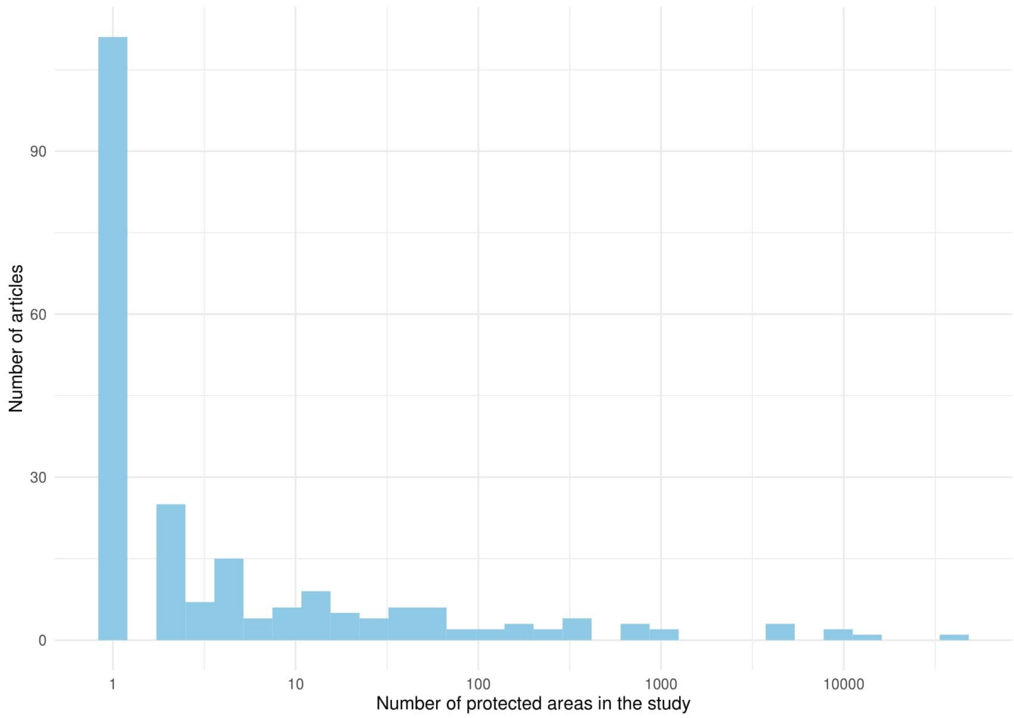

Figure 8. Number of protected areas studied in each study. X-axis is a log(10) scale.

The studies were predominantly conducted in the tropical-subtropical forest biome (122 studies), followed by temperate-boreal forests and woodlands (70) (Figure 9). Savannas and grasslands accounted for 68 studies, while deserts and semi-deserts were represented in 23 studies. Shrublands were examined in 40 studies, and alpine/polar regions in 22 studies. Freshwater ecosystems, including rivers and streams (25) and lakes (13), received less attention, as did wetland and coastal biomes, with palustrine wetlands (26) and brackish tidal zones (10) being less frequently studied. As certain studies examined multiple biomes, the total count exceeds the number of included studies.

Some sampling methods were far more prevalent than others. The most frequently used method was Geographical Information Systems (GIS), employed in 105 studies (Figure 9). Sampling plots were used in 58 studies, and stratified sampling in 50. Random sampling was applied in 34 studies, while transects accounted for 37 studies. Systematic sampling was reported in 42 studies, and point counts were used in 21 studies. Seven studies used third party databases. Finally, mark-recapture and GPS collars were the least frequently used methods, appearing in only 3 and 1 studies, respectively.

Most studies focused on subjects within the Plantae (170) and Animalia (110) Kingdoms. Few articles looked at Fungi, Protista, and Bacteria, each accounting for 3 studies.

The overwhelming majority of studies (279) were observational in nature, while experimental designs were much less common, appearing in 3 studies. Since some articles contained both observational and experimental components, the total number of studies exceeds the number of articles. In terms of experimental designs, most studies employed a combination of control-intervention (CI) and time-series designs (132). Ninety-nine studies used a CI design alone. Time-series designs made up 20 studies, while 23 combined time-series data with Before-After Control-Impact (BACI) designs. Pure BACI and Before-After (BA) designs were the least common, each represented by 4 studies

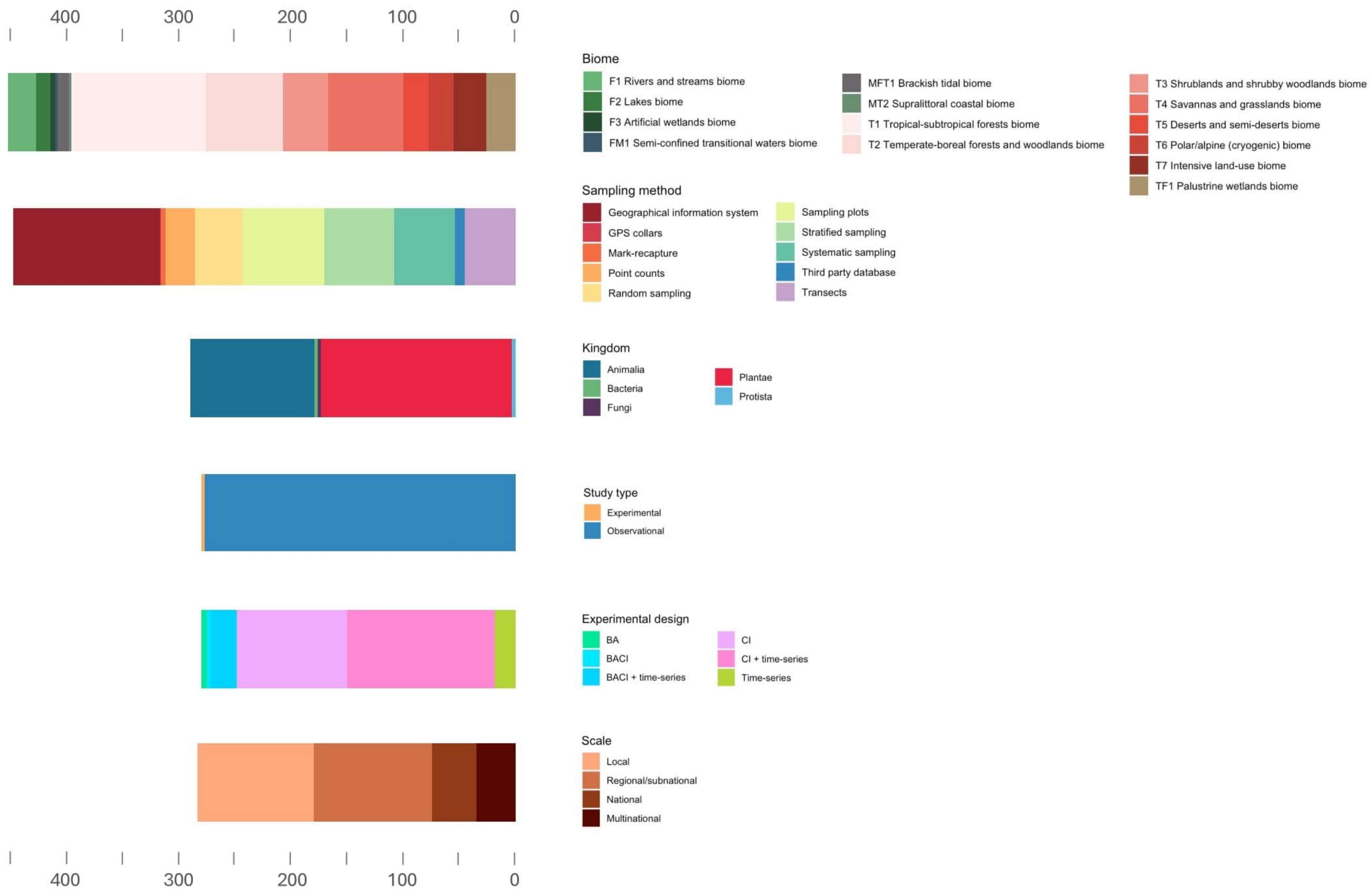

Figure 9. Studied biome, sampling method, kingdom, study type, experimental design, and scale across articles.

The most frequently used survey method was direct observation, employed in 129 studies (Figure 10). Satellite measurements were also commonly used, accounting for 105 studies. Indirect observation methods were used in 24 studies, while direct listening was applied in 16 studies. Questionnaire-based methods accounted for 4 studies, and camera traps were used in 11 studies. Hand searching was used in 15 studies, and net fishing in 11. Basket traps were used in 4 studies, live trapping in 7, sweep netting in 8, and DNA extraction in 5 studies. Less frequently applied techniques included pit-traps (2), kick sampling (2), active fishing (3), and spring-loaded bar traps (1).

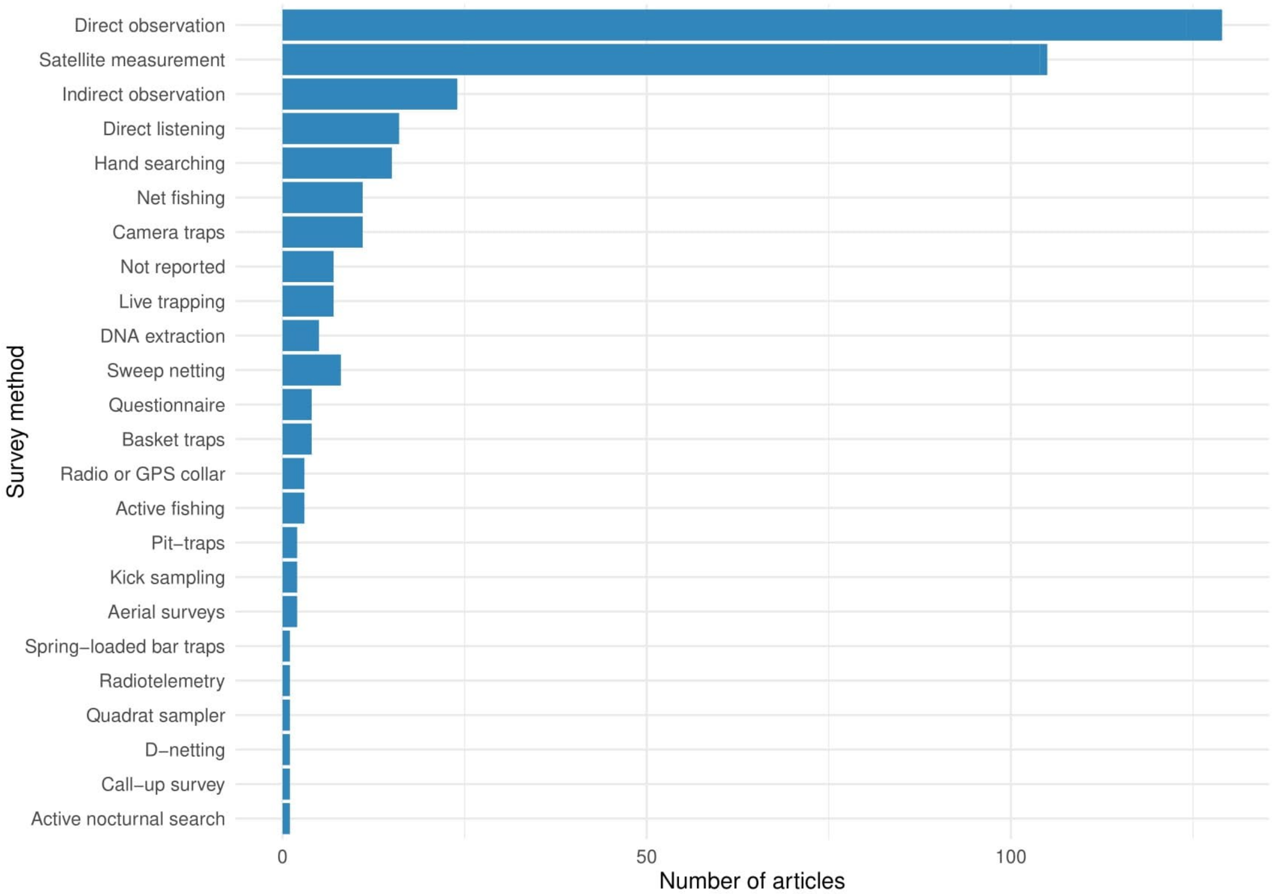

Figure 10. Number of articles by survey method used.

4.5. Temporal trends in study designs and sampling methods

The earliest studies had CI designs and those were dominant throughout time either with or without time series (Figure 11). The first study with a solely time series design is from 2001, but they don't increase in number much until 2014 onwards. The earliest study with BACI design appears in 2008. From 2014, studies with BACI design increased and were mostly implemented with a time series design.

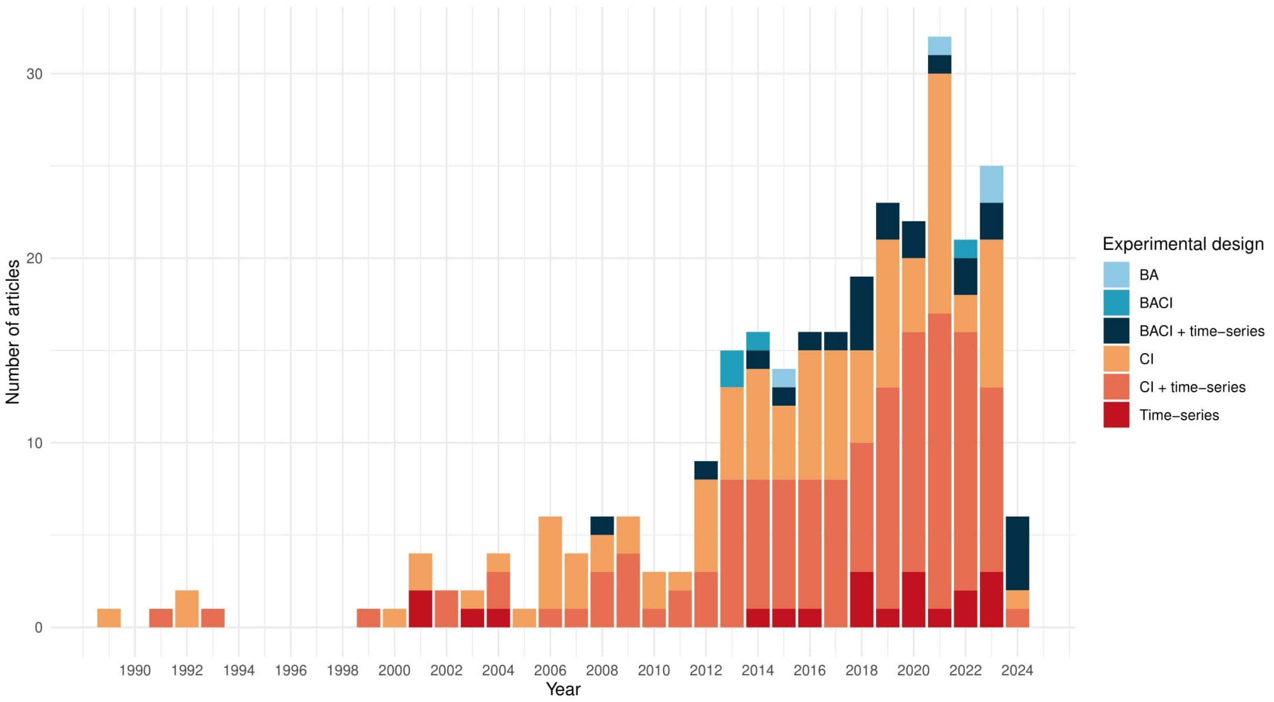

Figure 11. Number of articles across publication year by experimental design

Control-Intervention (CI) designs and CI combined with time-series data were the most frequently used across all study durations (Figure 12). Beyond one year, the number of studies decreases significantly, with only a few extending beyond 50 months. Longer-term studies predominantly used CI + time series designs, with other designs such as Before-After Control-Impact (BACI) + time series also present.

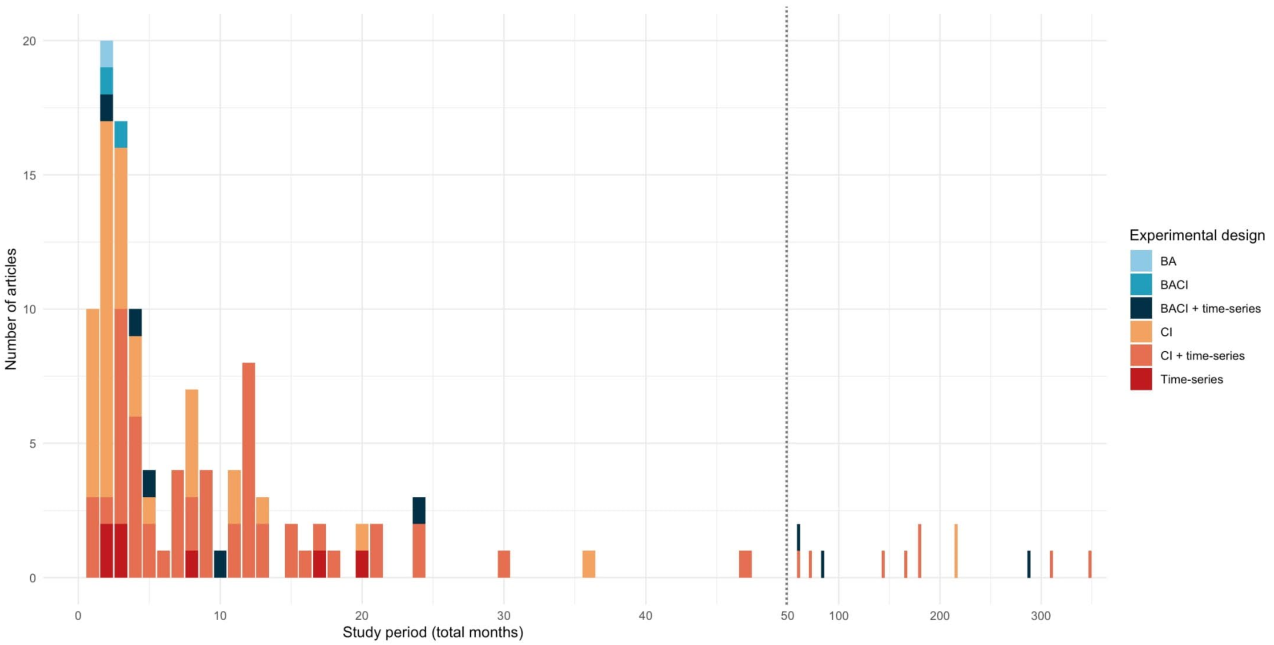

Figure 12. Articles using different study periods (total months) by experimental design. Note: Axis compression at a ratio of 2:1 above 50 months on the x-axis.

The variety of survey methods used increases over time with direct observation being a constantly used method from the earliest time (Figure 13). Studies using satellite measurement appeared after 2000 and their use increased after 2011. Studies using camera traps appear only after 2010. Similarly, studies employing direct listening are more common after 2014.

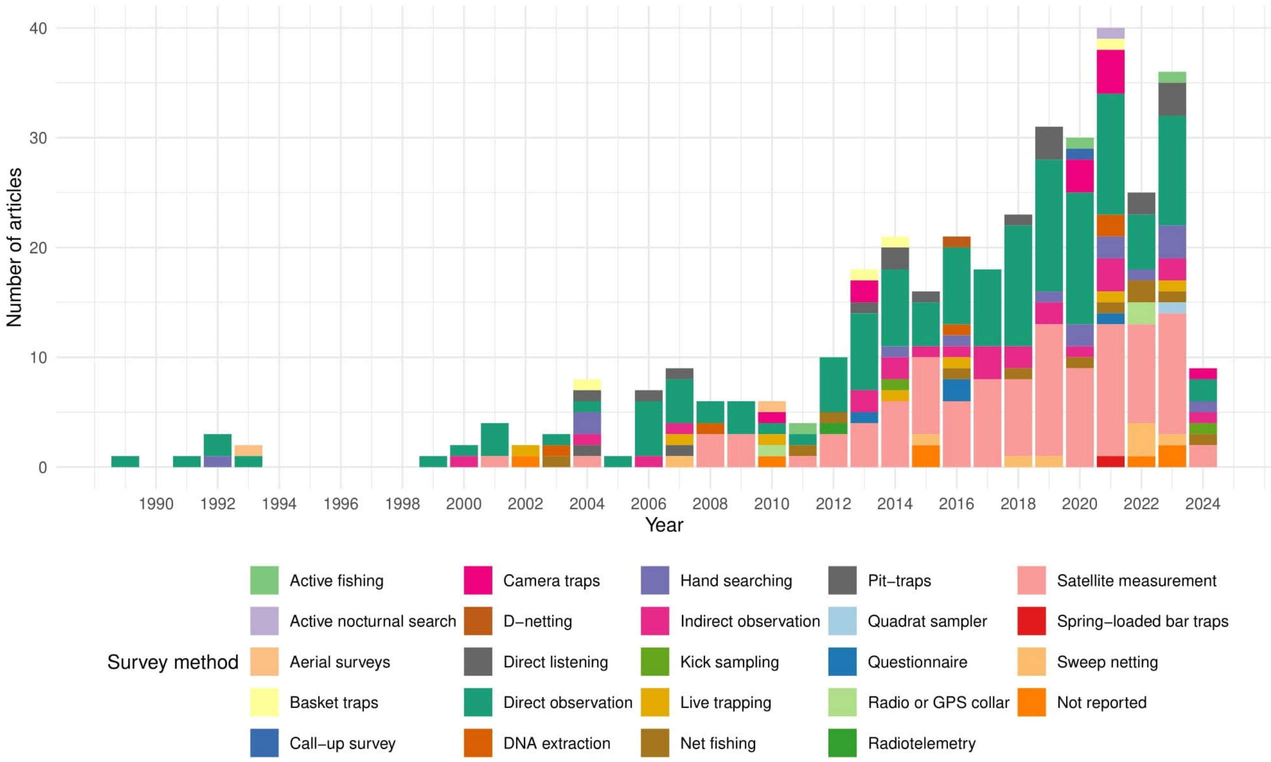

Figure 13. Number of articles across publication year by survey method

When considering total study duration across subjects, we observe longer-term studies to be more common for ecosystem and land-use studies, as well as for vegetation and woody plant research (Figure 14). A smaller number of long-term studies also focus on mammals and birds, though most studies on these taxa tend to be of shorter duration.

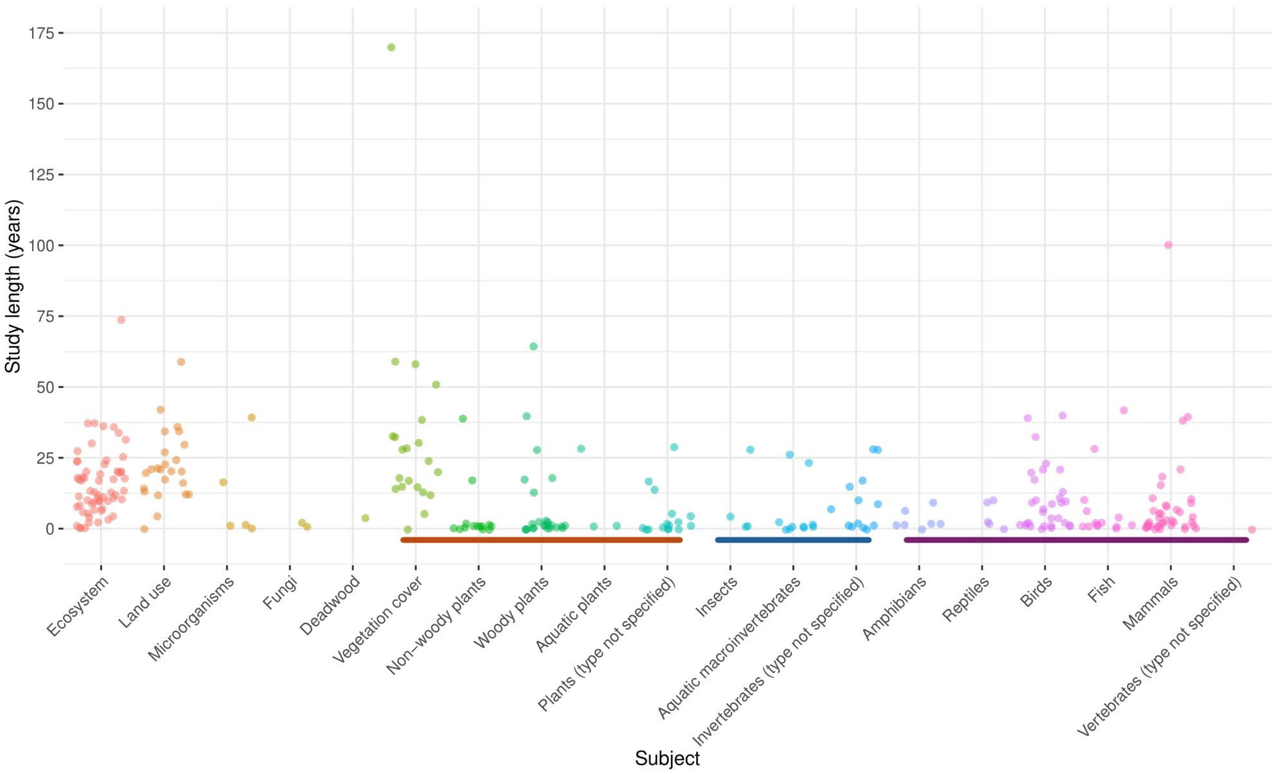

Figure 14. Study length by subject across included articles. Length is provided as full years (>12 months); thus, less than 12 months correspond to 0. Coloured bands in the x-axis indicate related but mutually exclusive categories: For these categories, the finest level provided is displayed. Orange = vegetation groups. Blue = invertebrate groups. Purple = vertebrate groups.

The Essential Biodiversity Variables (EBVs) investigated in the included articles varied considerably, with a strong emphasis on ecosystem and population-level metrics (Table 5). Species abundance was the most frequently studied EBV (119 studies), followed by community abundance (116) and live cover fraction (91). Ecosystem disturbance (67), species distributions (61), and ecosystem distributions (56) were also commonly assessed. In contrast, genetic composition EBVs, such as genetic differentiation (3) and genetic diversity (9), were underrepresented, along with interaction diversity (6) under community composition. Similarly, species traits such as movement (17), phenology (3) physiology (4), and reproduction (11) received relatively little attention. Ecosystem metrics, including ecosystem vertical profile (8) and primary productivity (5) were also rarely investigated.

Table 5. Number of articles investigating different Essential Biodiversity Variables (EBVs)

EBV group	EBV	n
Genetic composition	Genetic diversity (richness and heterozygosity) / Intraspecific genetic diversity	9
	Genetic differentiation (number of genetic units and genetic distance)	3
	Effective population size	0
	Inbreeding	0
Species populations	Species distributions	61

	Species abundances	119
Species traits	Morphology	36
	Physiology	4
	Phenology	3
	Movement	17
	Reproduction	11
Community composition	Community abundance	116
	Taxonomic/phylogenetic diversity	24
	Trait diversity	27
	Interaction diversity	6
Ecosystem functioning	Primary productivity	5
	Ecosystem phenology	0
	Ecosystem disturbances	67
Ecosystem structure	Live cover fraction	91
	Ecosystem distribution	56
	Ecosystem Vertical Profile	8

4.6. Cross-variable patterns in the evidence base

Community abundance and species abundance were the most frequently studied EBVs across a wide range of survey methods (Figure 15). Community abundance was most commonly studied using direct observation (87), direct listening (14), net fishing and sweep netting (8 each). Similarly, species abundances were predominantly assessed using direct observation (56), direct listening (14) and indirect observation (14). Live cover fraction, another frequently studied EBV, was strongly associated with satellite remote sensing (67) and direct observation (28). Ecosystem disturbances and distributions showed similar patterns, with 60 and 52 studies using satellite remote sensing, and 11 and 8 using direct observation, respectively. EBVs related to genetic and species traits were sparsely studied across all survey methods. The intersections also reveal specific gaps in using certain methods for specific EBVs. For instance, interaction diversity was rarely studied, with limited reliance on indirect observation (5), hand searching (4), live trapping (1) and net fishing (1).

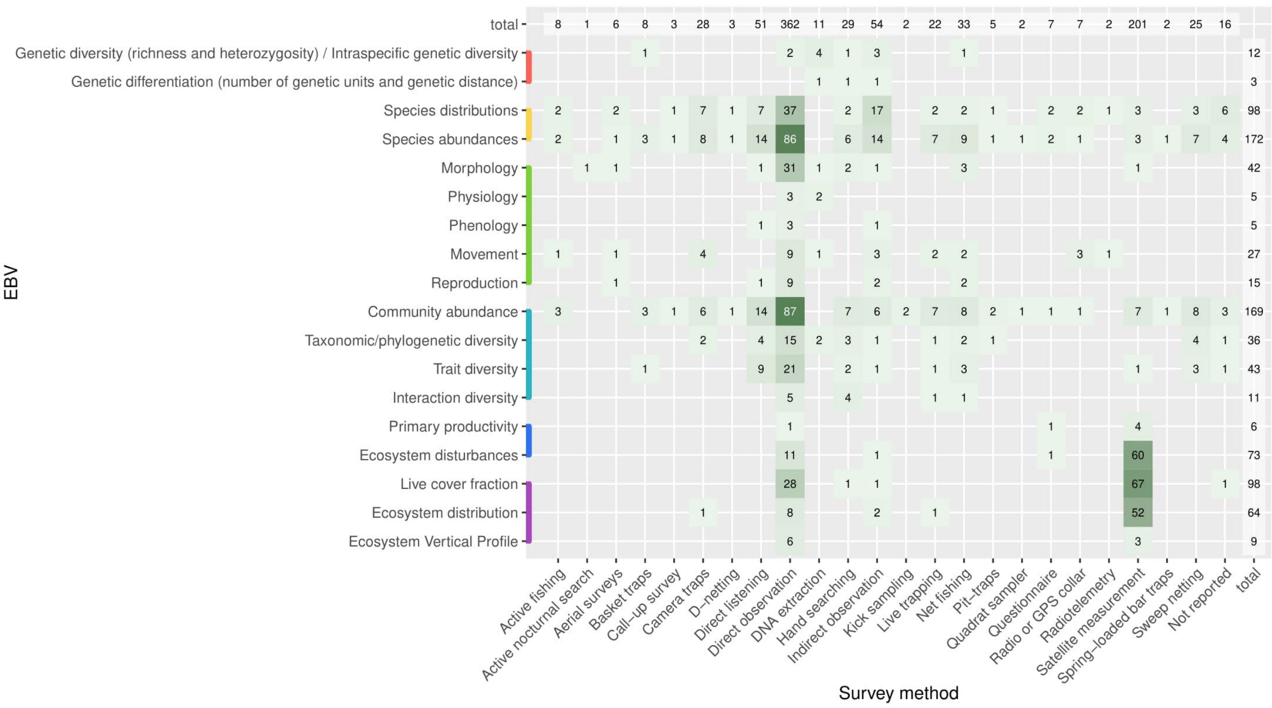

Figure 15. Essential Biodiversity Variables (EBVs) examined by survey method in included articles. Red = Community composition, yellow = Ecosystem functioning, green = Ecosystem structure, turquoise = Genetic composition, Blue = Species populations, purple = Species traits.

Considering the intersections between EBVs and sampling methods, live cover fraction and ecosystem disturbances were primarily measured using GIS, with 67 and 60 instances, respectively (Figure 16). Species abundances were frequently assessed using sampling plots (40), stratified sampling (31), and systematic sampling (30). Species distributions also showed strong association with systematic sampling (17), transects, and sampling plots (13 each), while community-level metrics, such as community abundance, were heavily reliant on sampling plots (44), systematic sampling (28), and stratified sampling (27). In contrast, EBVs related to genetic and trait diversity, such as genetic diversity and morphology, were less represented across all sampling methods. For instance, genetic diversity was most linked to systematic sampling, with only 4 studies. The intersections also highlight gaps in using specific methods for certain EBVs. For example, interaction diversity, primary productivity, and ecosystem vertical profile were minimally assessed, regardless of method. Notably, specialised methods such as GPS collars and mark-recapture were rarely applied, suggesting limited attention to movement and behavioural EBVs.

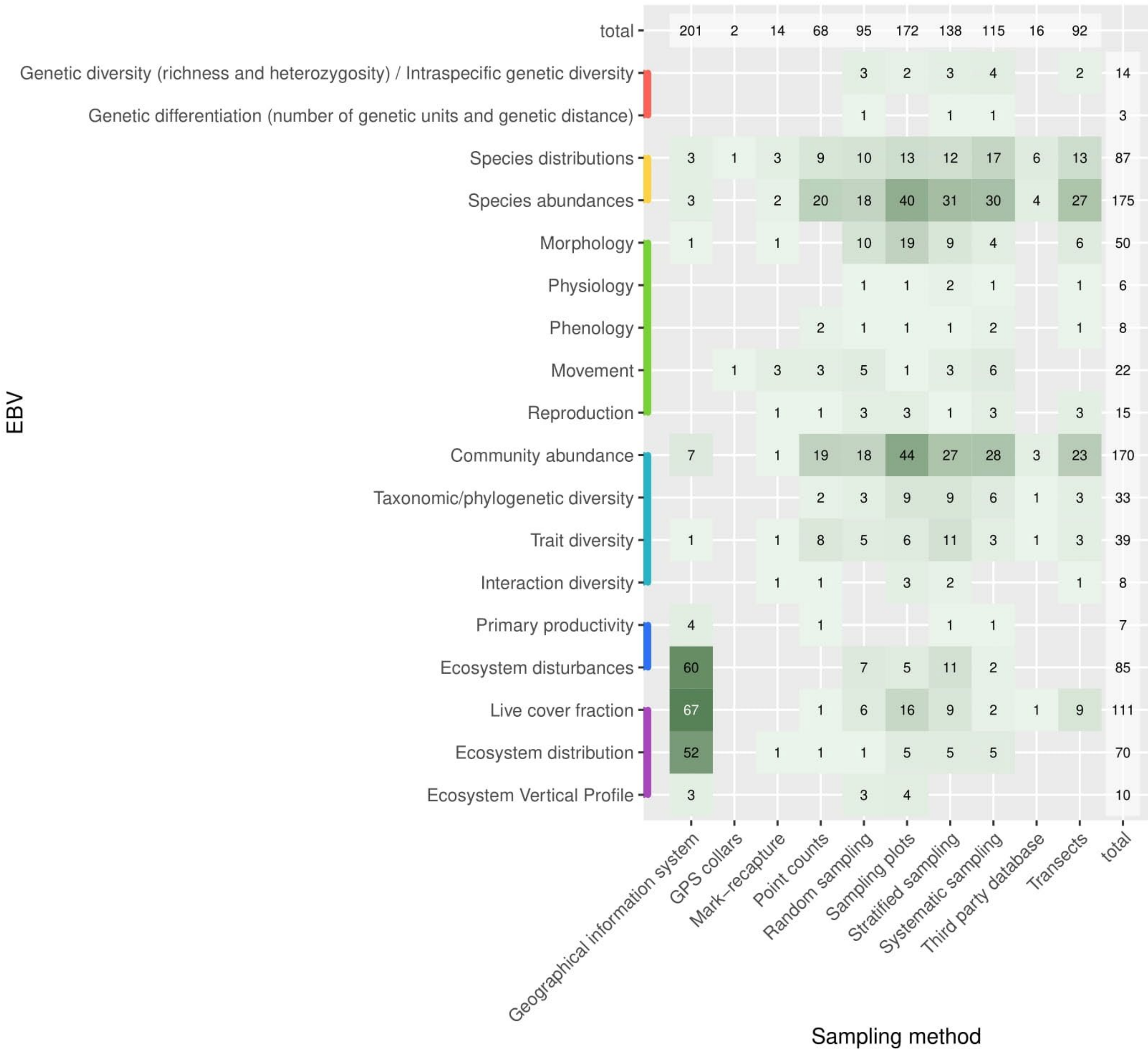

Figure 16. Essential Biodiversity Variables (EBVs) examined by sampling method in included articles. Red = Community composition, yellow = Ecosystem functioning, green = Ecosystem structure, turquoise = Genetic composition, Blue = Species populations, purple = Species traits.

Experimental designs across EBVs show species and community abundance assessed using CI to be the most frequent study types, each with 57 studies (Figure 17). This was followed by CI + time series (49 and 48 studies, respectively). Similarly, live cover fraction was frequently evaluated using CI + time series (41), along with ecosystem and species distribution (30 and 32 studies, respectively). BACI and BACI + time series remains underutilised across most EBVs, with ecosystem-level metrics being the most represented group. EBVs such as phenology, movement, and physiology were rarely studied across any experimental design, reflecting gaps in evaluating finer-scale ecological processes.

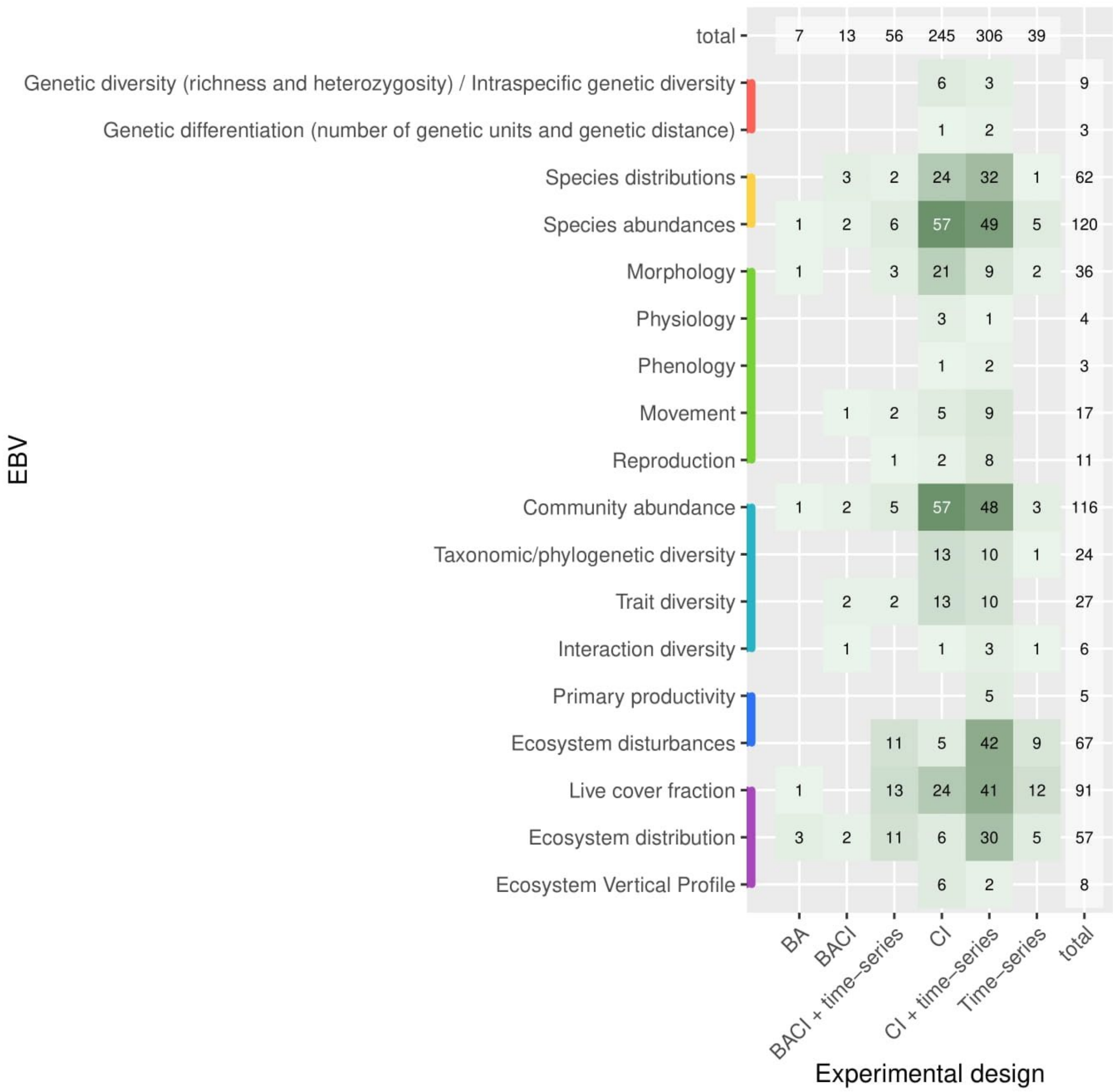

Figure 17. Essential Biodiversity Variables (EBVs) examined by experimental design in included articles. Red = Community composition, yellow = Ecosystem functioning, green = Ecosystem structure, turquoise = Genetic composition, Blue = Species populations, purple = Species traits.

Studies focusing on forest cover and forest cover loss (under Ecosystem) on tropical and subtropical forests were the most common (43) (Figure 18). Studies on mammals within the same biome were the second most common (21). Plants were more commonly studied in the savannas and grasslands biome as well as the forest biomes from tropical to boreal forests. Temperate-boreal forest and savannas and grassland biomes had the highest numbers of studies on birds (15 and 12 respectively) and invertebrates (7 in both). Studies on land use focused on tropical-subtropical forest biome (13) and intensive land-use biome (e.g., agricultural lands) (9).

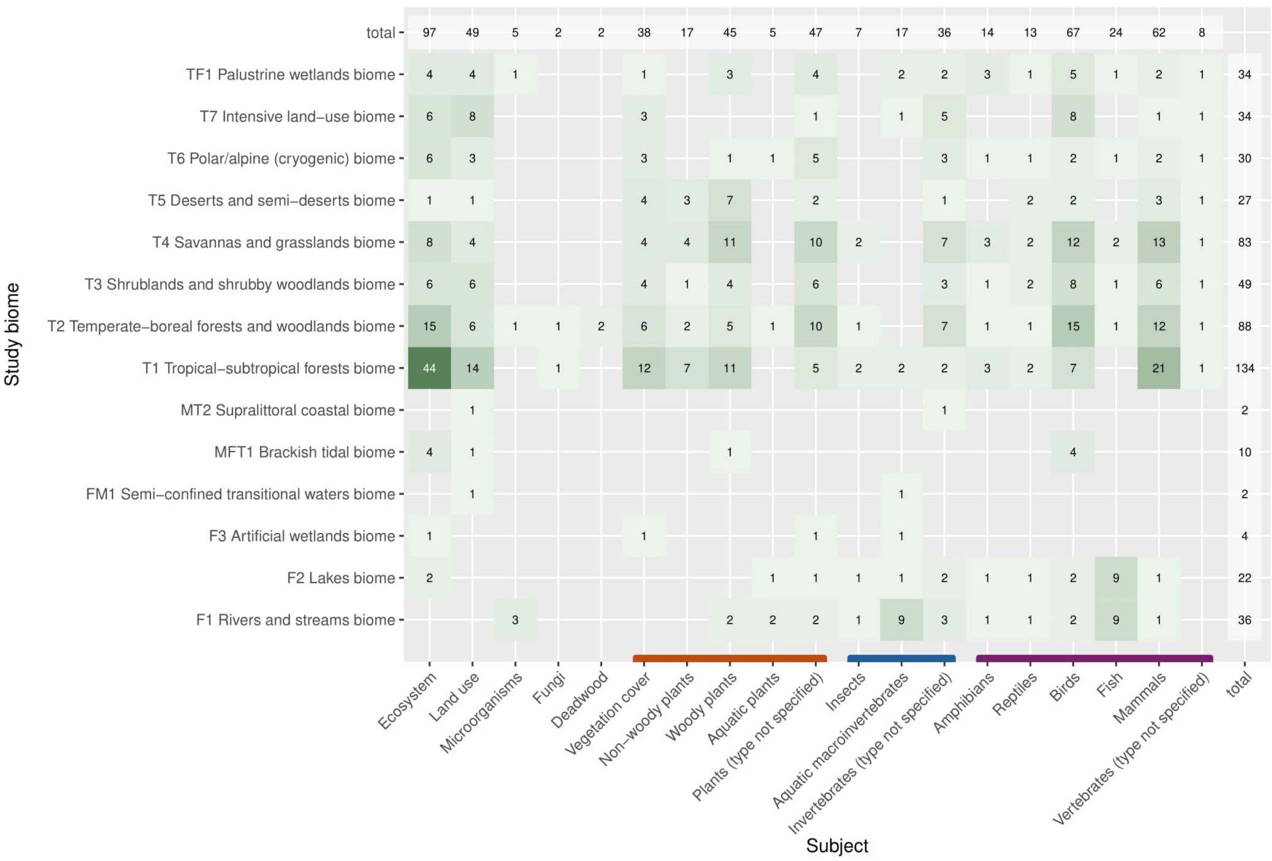

Figure 18. Study biomes by subject, showing the number of articles including both. Coloured bands in the x-axis indicate related but mutually exclusive categories: For these categories, the finest level provided is displayed. Orange = vegetation groups. Blue = invertebrate groups. Purple = vertebrate groups.

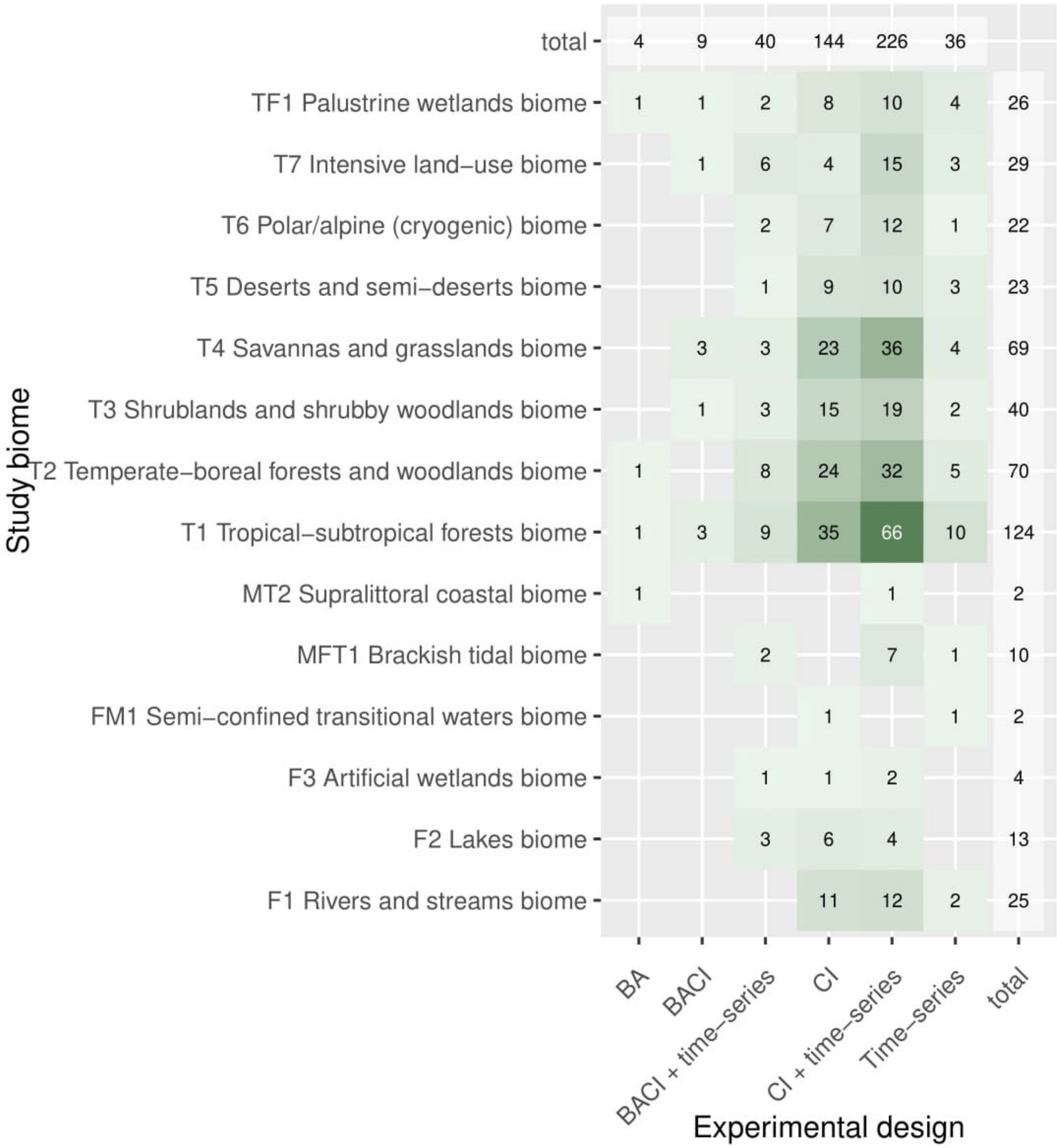

Figure 19. Study biomes and experimental designs used across articles.

The highest number of studies with CI + timeseries design (66) were conducted in the tropical-subtropical forest biome followed by savannas and grasslands biome (36) and temperate-boreal forest biome (32) (Figure 19). Studies with BACI + timeseries design were most common in the tropical-subtropical and temperate-boreal forest biomes (9 and 8 studies respectively) followed by intensive land use biome (6).

Across all sampling methods, Africa had the highest number of studies, except for those employing Geographical Information System (GIS) methods and third-party databases (Figure 20). GIS-based methods were the most frequently used, and most applied in Asia (25), Africa (19), followed by Central and South America (16 each). GIS, point counts, and stratified sampling were the most widely represented methods across continents, although none was represented in all continents. The highest number of studies using sampling plots were conducted in Africa (22), followed by Europe (13) and South America (7). Stratified sampling was primarily conducted in Africa (12), Asia (10) and Europe (8). Transect-based methods were mostly used in Africa (11), with fewer studies conducted in Asia (6) and Europe (6).

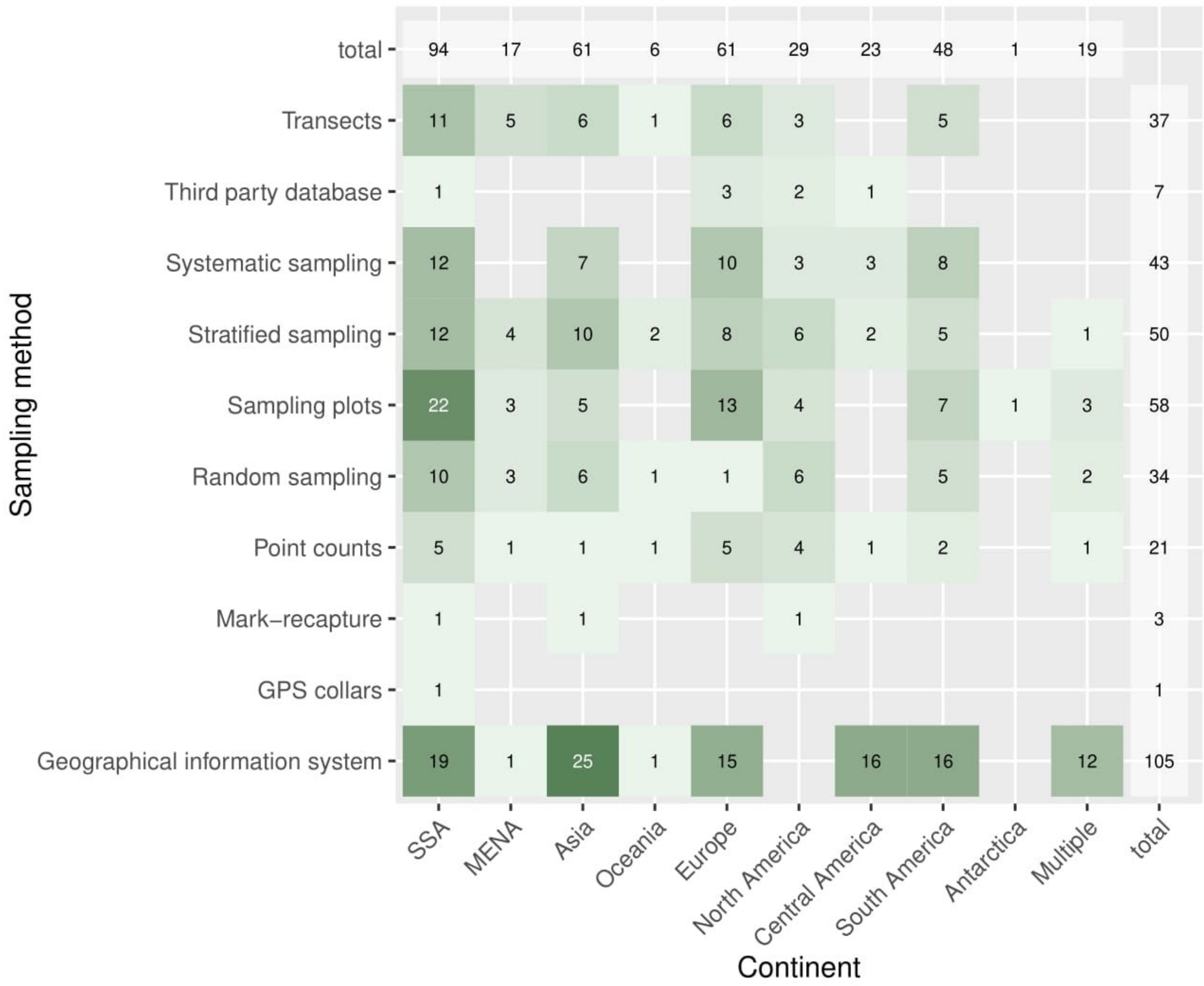

Figure 20. Sampling methods used across continents in included articles. SSA = Sub-Saharan Africa. MENA = Middle East and North Africa. Multiple = more than one continent.

GIS methods were the most frequently used across all experimental designs, particularly in studies employing Control-Intervention (CI) designs combined with time series (57), followed by BACI + time series studies (17) (Figure 21). Sampling plots were also widely applied, with the highest numbers observed in CI and CI + time series (30 and 23, respectively). CI and CI + time series were the most frequent designs across all sampling methods, with CI + time series being the only design represented across all sampling methods. GPS collars was the least utilised method, and along with mark-recapture were the only methods without CI studies. Only four studies used a Before-After (BA) design, and were based on GIS (2), sampling plots (1), and transects (1). The majority of BACI + time-series utilised GIS (17), followed by point counts (4), stratified sampling (3), and sampling plots (2). The remaining studies employed transects, random sampling, and mark-recapture (1 study each). In contrast, BACI studies without a time-series component used sampling plots (2), GIS, point counts, mark-recapture, and transects (1 study each).

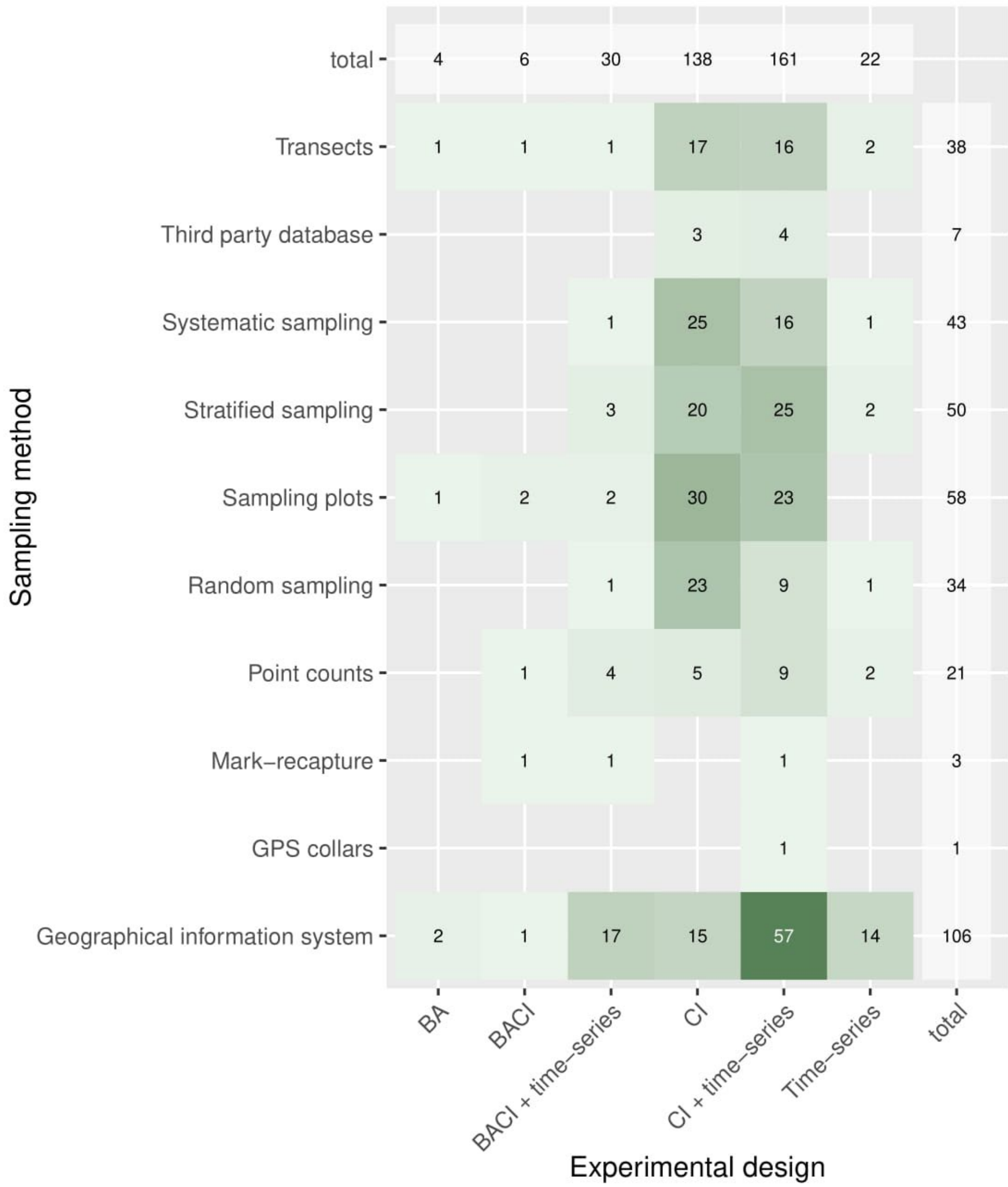

Figure 21. Sampling methods and experimental designs used across included articles.

Satellite measurements were the most applied survey method, particularly for ecosystems (56 studies) and land use (23), followed by vegetation (21) (Figure 22). Direct observation was the only survey method represented in all recorded subjects, with the highest application in woody plants (28), birds (28), and mammals (19). Indirect observation was predominantly used for mammals (21). Some methods, such as kick sampling, quadrat sampling, and pit-trapping were rarely used.

Figure 22. Survey methods and subjects used across the included articles. Coloured bands in the x-axis indicate related but mutually exclusive categories: For these categories, the finest level provided is displayed. Orange = vegetation groups. Blue = invertebrate groups. Purple = vertebrate groups.

Studies conducted at the regional or subnational scale were the most common across many subjects, with notable emphasis on ecosystems (18) and woody plants (17), and vegetation (12) (Figure 23). Local-scale studies were also well represented, particularly for mammals (23), birds (14), woody plants (14), and ecosystems (15). National-scale studies were less frequent overall, but still represented across ecosystems (17), mammals (7), and birds (6). Multinational studies focused mainly on ecosystems (14) and land use (8), followed by birds (5). Overall, the scale of study tended to align with the subject, with local and regional studies being most common for field-based and species-focused research, while multinational studies were more prevalent for broad-scale ecological topics like ecosystems and land use.

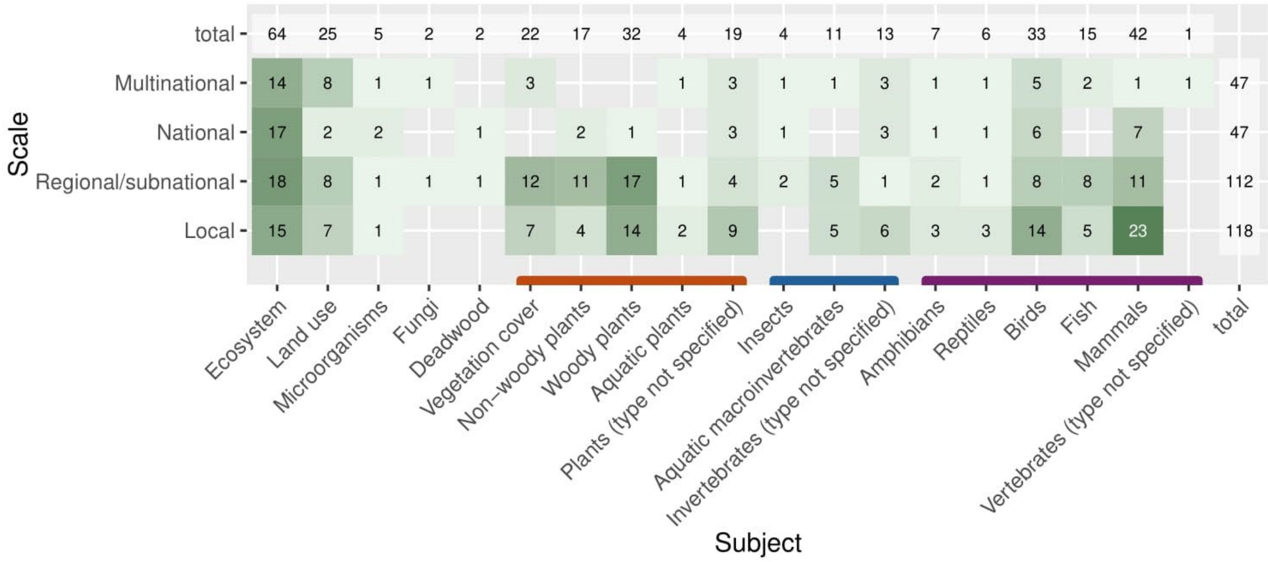

Figure 23. Study scale and subject across included articles.

CI and CI + time-series designs were the most frequently applied experimental designs across all subjects (Figure 24). CI + time-series was especially prominent for ecosystems (34), land use (15), and mammals (25). CI alone was widely used in studies on woody plants (20) and mammals (15). BACI + time-series studies were relatively limited but still had notable representation in land use (5), ecosystems (8), and woody plants (3). Time-series designs alone were less common overall but were employed in ecosystems (8) and vegetation (5). BA and BACI designs were rarely used, with only a few studies focused on land use, ecosystems, invertebrates and mammals. Mammals and birds stood out among vertebrates, with CI and CI + time-series being the dominant designs (15 and 25 studies for mammals; 8 and 17 studies for birds).

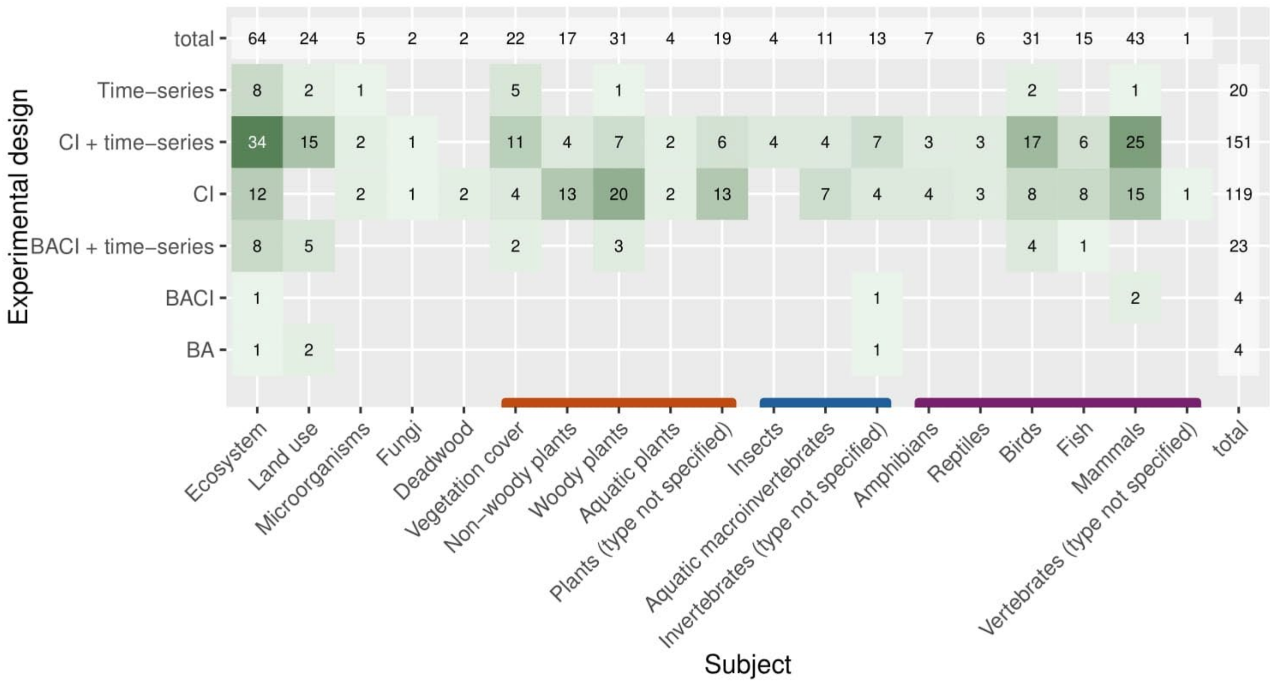

Figure 24. Experimental design and subject used across the included articles.

Asia led in number of studies on ecosystems (17), land use (7), and mammals (7). Europe showed notable representation in birds (9), ecosystems (9), and invertebrates (8) (Figure 25). Africa had significant representation in woody plants (13), mammals (12), and ecosystems (8). Central and South America were well-represented for ecosystems (9 studies each), with South America showing additional focus on vegetation (6), mammals (8), and fish (7 studies). North America was less represented overall but included studies on birds (6) and plants (4). MENA had most representation in non-woody (6) and woody (5) plants, while Oceania had limited representation, with just two studies on mammals and single studies on fish, birds, aquatic macroinvertebrates, and ecosystems. Antarctica had no representation for most subjects except for a single study on aquatic plants. Studies considering multiple continents had modest representation, primarily focusing on ecosystems (9) and land use (3). Overall, mammals, ecosystems, and birds were the most studied subjects, with significant contributions from Asia, Europe, and Africa, whereas microorganisms, fungi, and deadwood were sparsely studied across all continents.

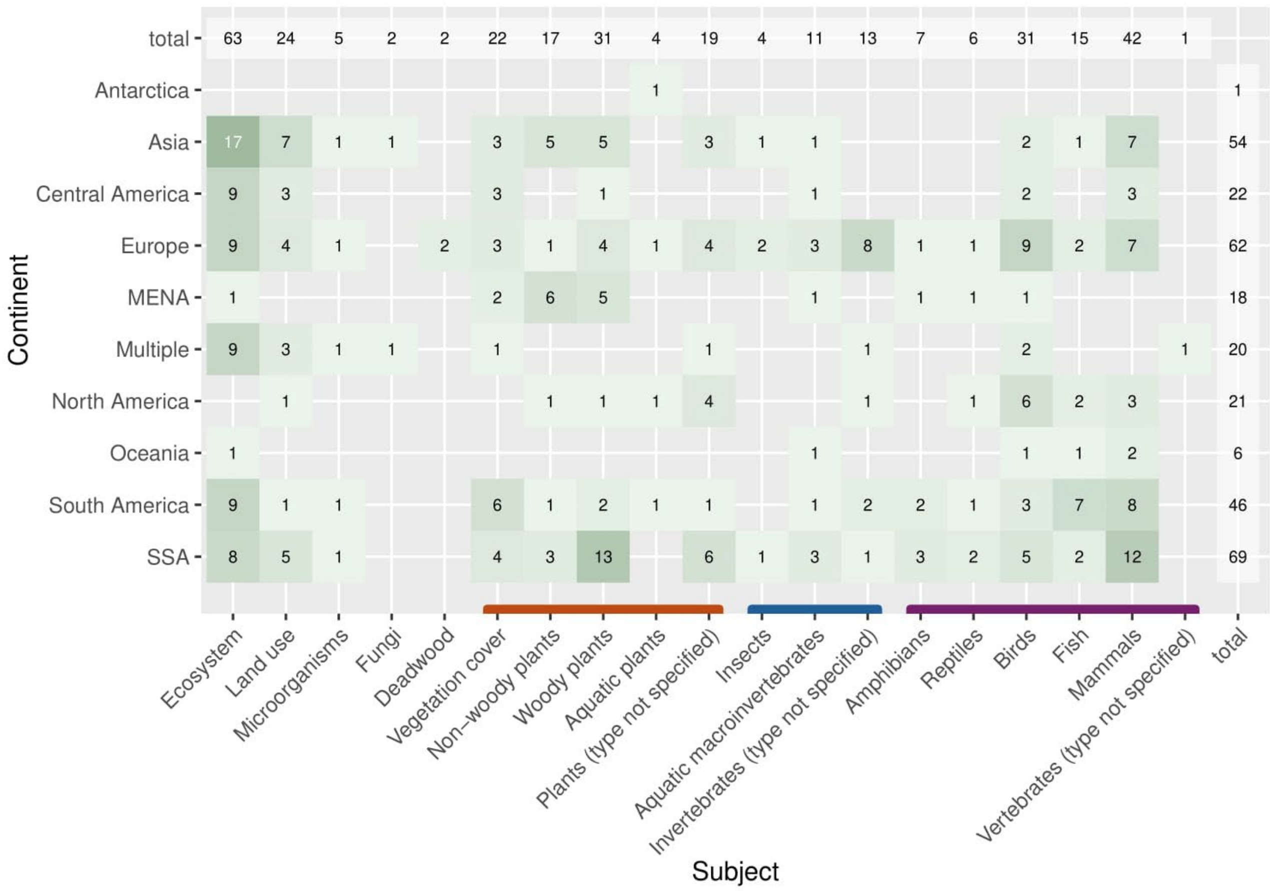

Figure 25. Subjects studied by continent across included articles. Coloured bands in the x-axis indicate related but mutually exclusive categories: For these categories, the finest level provided is displayed. Orange = vegetation groups. Blue = invertebrate groups. Purple = vertebrate groups.

The intersections between survey and sampling methods reveal both expected trends and underexplored areas in biodiversity monitoring. Unsurprisingly, satellite measurement was overwhelmingly associated with Geographical Information Systems (GIS) (105 studies). Similarly, direct observation was frequently paired with field-based sampling methods such as sampling plots (52), stratified sampling (32), and systematic sampling (31), which are well-established approaches for species and population assessments. More specialised techniques, such as DNA extraction, and radio or GPS collars, showed limited use and were primarily linked to systematic sampling, indicating their more targeted application in genetic and movement-based studies. Indirect observation demonstrated broader versatility, being applied across all but three survey methods. Methods such as kick sampling, pit-traps, and spring-loaded bar traps were minimally represented.

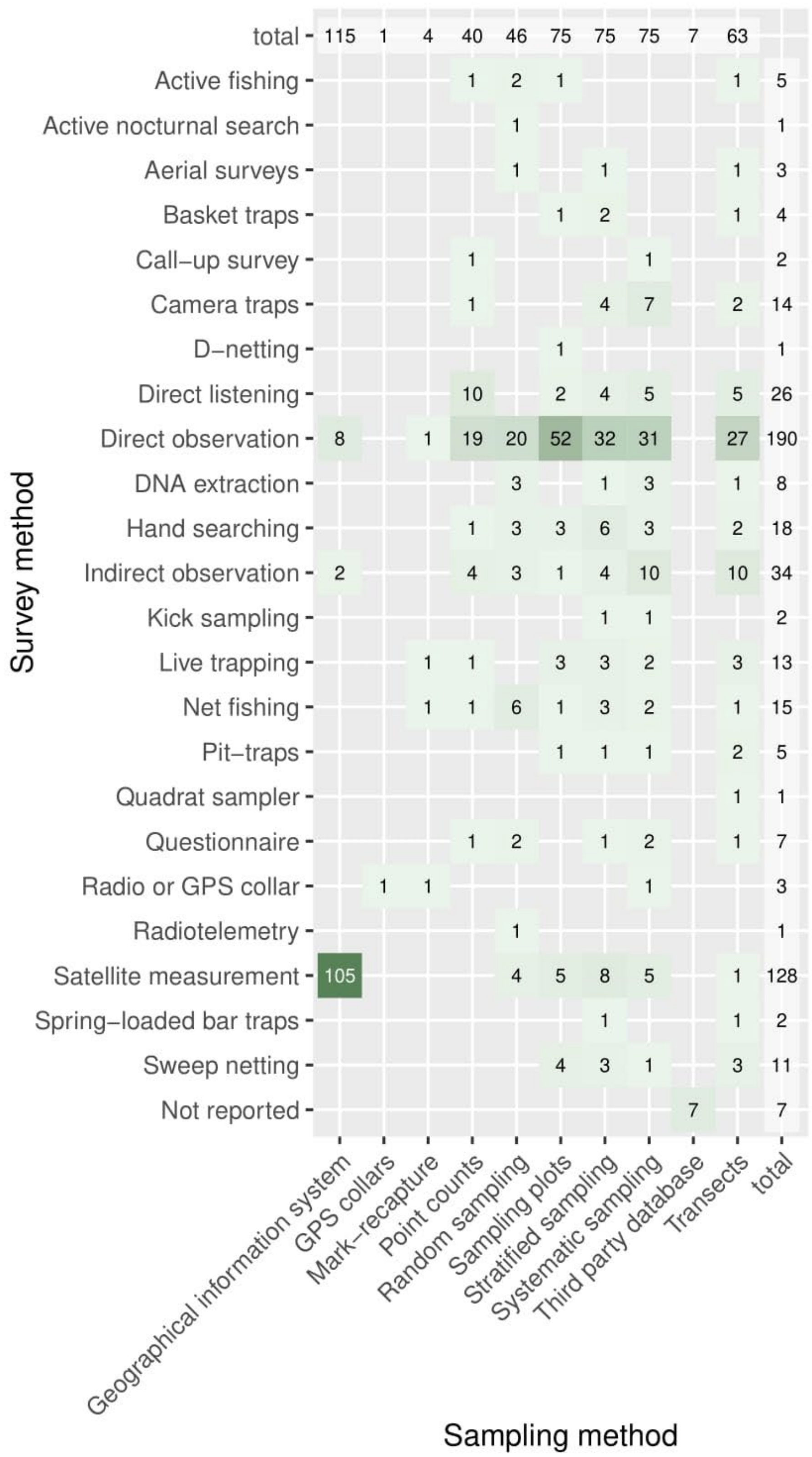

Figure 26. Survey method by sampling method for included articles.

IUCN category II protected areas were the most frequently studied across all sampling methods, and amongst all IUCN categories, particularly with GIS (44), sampling plots (34), stratified sampling (24), transects (25), and systematic sampling (27) (Figure 27). Overall, GIS was the most widely used method across all PA designations except for UNESCO World Heritage Sites where stratified sampling was more frequent (6 studies, versus 5 using GIS). Studies considering IUCN Categories Ia, Ib, III and IV showed similar patterns, with most relying on GIS, followed by stratified sampling and sampling plots in comparable proportions. Fewer studies looked at IUCN categories V and VI protected areas, considered as multiple-use areas. For these categories, 19 and 14 studies used GIS, respectively, while stratified sampling (5 and 3) and sampling plots (7 and 3) were less commonly applied. Natural 2000 studies predominantly used GIS (5) and sampling plots (5). Studies on UNESCO World Heritage Sites and Biosphere Reserves were represented in similar proportions and employed a comparable range of sampling methods across categories. Studies categorised under Other (e.g., reserves in countries that do not follow IUCN classifications, such as China or South Africa, certain provincial parks, forest reserves, etc.) and Not reported frequently employed GIS (12 and 24, respectively). Other reserves also frequently used sampling plots (14) and stratified sampling (15).

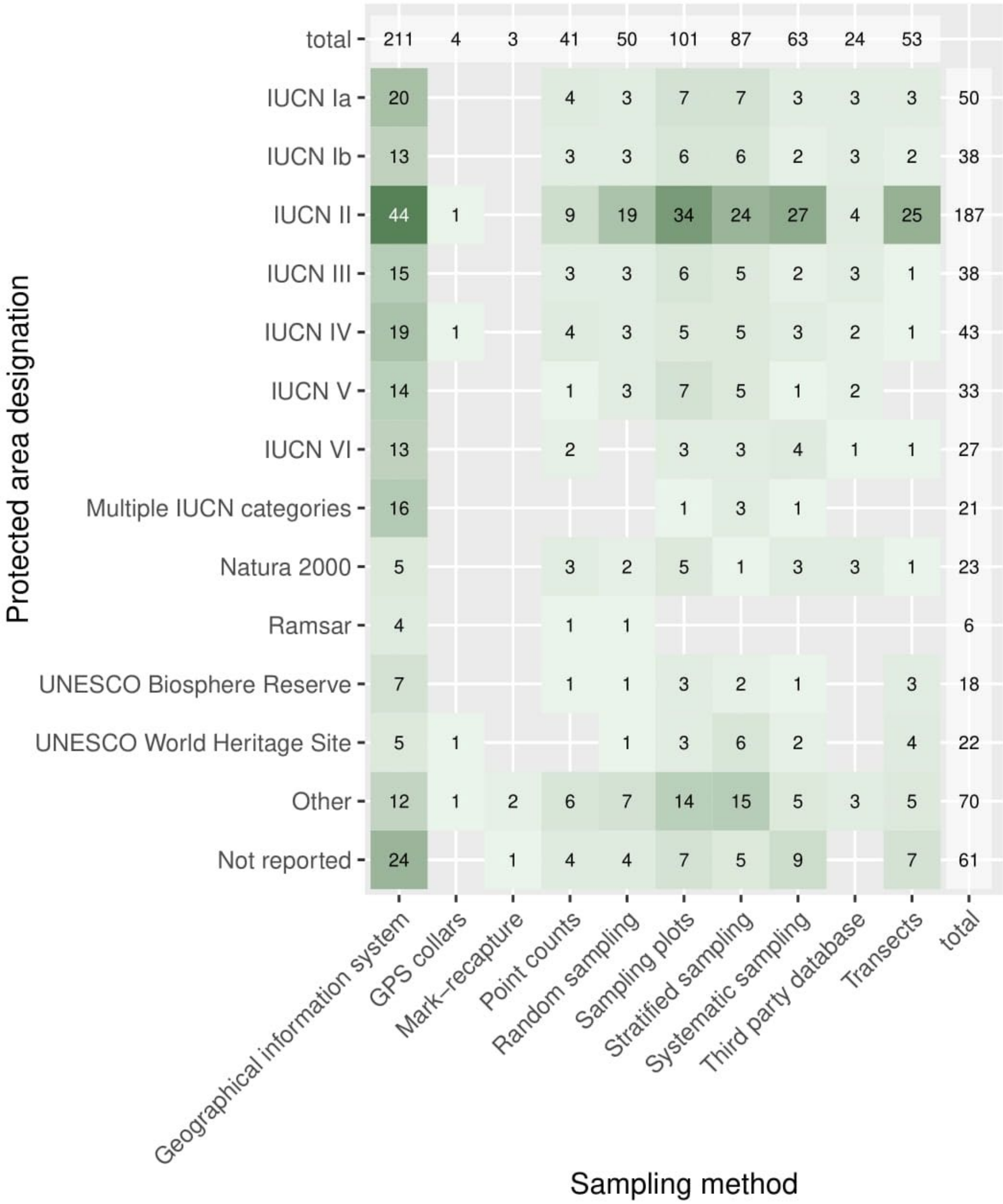

Figure 27. Sampling methods used across protected area designations in included articles. *Other* refers to protected areas that do not follow standard IUCN classifications, including reserves in countries such as China or South Africa, certain provincial parks, forest reserves, and other regionally or nationally designated conservation areas.

Studies on IUCN Category II protected areas predominantly relied on direct observation (65) and satellite measurements (44), the most used survey methods overall (Figure 28). Categories Ia-Ib followed a similar pattern, as did Natura 2000 sites, which primarily used satellite measurements (10) and direct observation (5). Ramsar sites, UNESCO Biosphere Reserves, and World Heritage Sites were less frequently studied but also relied on these methods. Among the less frequently studied categories, such as IUCN V and VI, satellite measurements and GIS remained the primary approaches. Studies categorised as Other or Not reported showed similar trends, with a strong emphasis on satellite data and direct observation.

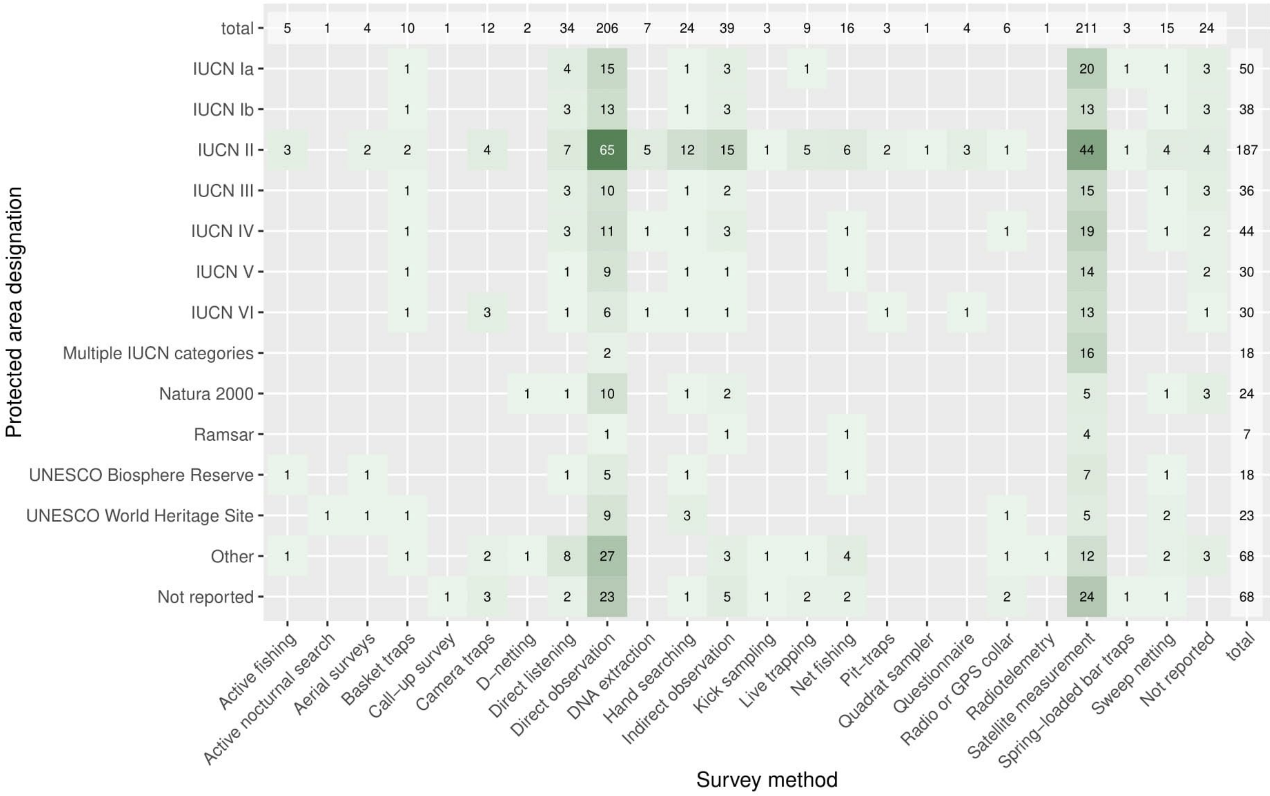

Figure 28. Survey methods used across protected area designations in included articles. *Other* refers to protected areas that do not follow standard IUCN classifications, including reserves in countries such as China or South Africa, certain provincial parks, forest reserves, and other regionally or nationally designated conservation areas.

5. Discussion and Conclusions

Protected areas are vital in addressing global biodiversity challenges, yet their effectiveness in achieving conservation outcomes remains a topic of considerable debate. Assessing PA effectiveness requires understanding whether they are located in areas that inherently support higher biodiversity or whether they actively contribute to conservation outcomes. This review focuses on the designation effect, evaluating whether biodiversity differs between protected and unprotected areas (Li et al., 2024).

Through this systematic map, we sought to collate and synthesise global evidence on the methods and metrics used to assess the effectiveness of terrestrial PAs in conserving biodiversity. By doing so, we aimed to provide a comprehensive overview of the existing evidence base, identifying patterns, gaps, and areas for methodological improvement.

5.1. Rise of monitoring and different methods

The rise of different monitoring methods from 2000 onwards corresponds with the political developments to halt biodiversity loss. The early international commitment to halt biodiversity loss by 2010 led to the need to examine and monitor progress. The first set of global biodiversity indicators were adopted in 2004 at the 7th UN Convention on Biological Diversity (CBD) and formed the basis for the first biodiversity indicators adopted by the EU Environment Council later in 2004 (EEA 2007). The sharp increase in studies since 2012 coincides with global conservation initiatives such as the Aichi Biodiversity Targets, which were adopted in 2010 under the CBD. These targets set ambitious goals for biodiversity conservation, including expanding protected area coverage and improving the effectiveness of conservation measures (cf. Aichi Target 11). The implementation of these targets likely spurred increased research efforts to monitor progress and evaluate conservation outcomes (Visconti et al., 2019).

The evidence-base reflects a strong reliance on certain well-established methods and metrics. Direct observation and satellite remote sensing emerged as the most frequently employed methods, reflecting their accessibility and broad applicability in conservation research. Direct observation is, and has traditionally been, a core method for a fine-scale, local level information whereas satellite remote sensing represents the other extreme of scale focusing on ecosystem-level structure and function (Proença et al., 2017). There is a rise in studies using satellite remote sensing since 2008 when mapping of forest cover and changes in it became more common (e.g. Hansen et al., 2010, Hansen et al., 2013, Potapov et al., 2008).

Techniques like mark-recapture and pit traps were rarely used, indicating a potential gap in research focused on species-level processes. This may be due to the resource-intensive nature of these methods, as they require significant effort, time, and specialised expertise, limiting their applicability for broad-scale assessments of PAs. As these methods also tend to focus on specific taxa, this may not align with the objectives of many PA effectiveness studies. Similarly, the limited genetic composition and species trait metrics may be due to the specialised expertise and significant resources required for such studies. However, conducting species-level assessment is important as proxy indicators such as forest or vegetation cover may not capture species-level processes, because these can be non-linear (de Araujo Martins et al., 2023). This can lead to a situation where changes in species abundance, richness and/or composition go unnoticed.

5.2. Control-intervention studies dominant in study designs

The dominance of Control-Intervention (CI) designs, with or without time-series data, also underscores a reliance on spatial comparisons to assess PA effectiveness. Without a temporal reference, CI designs may struggle to distinguish designation effects from management effects. While these approaches can provide valuable insights into ecosystem and population-level changes, they may overlook finer-scale processes, and risk producing inaccurate or biased estimates of biodiversity responses. Additionally, studies evaluating the complementarity of PA networks, i.e., whether biodiversity is higher inside the network than outside, should be distinguished from those assessing PA effectiveness, even if they are often labelled as such.

Experimental designs such as Before-After Control-Impact (BACI) were underrepresented, though their adoption has increased in recent years corresponding to the general increase in awareness of the importance of impact evaluation in conservation research. The underrepresentation noticed here is likely due to the inherent challenges of implementing BACI designs, especially in the context of PAs, as such studies require baseline data collected prior to the establishment of a protected area, which is often unavailable due to historical lack of monitoring efforts (Christie et al., 2019). Additionally, identifying appropriate control sites that are ecologically comparable can be challenging, particularly in highly modified or fragmented landscapes.

Despite these challenges, BACI designs often offer advantages in assessing PA effectiveness. By incorporating a temporal element and control sites, BACI studies provide stronger causal inference than simple spatial comparisons, reducing bias associated with pre-existing ecological differences between protected and unprotected areas (Christie et al., 2021). This makes them particularly valuable for detecting changes in biodiversity and ecosystem processes over time and, importantly, for assessing the causal relationship with PA creation. However, the feasibility and robustness of BACI studies often hinges on early planning and integration of effectiveness evaluation into the design phase (Smokorowski & Randall, 2017). If monitoring efforts are only introduced post hoc, the opportunity to apply BACI rigorously is often lost. It should also be noted that in some cases, counterfactual assessments may be less relevant for management when conservation actions, like strict law enforcement, are clearly linked to species recovery, and alternative scenarios are difficult to establish due to species being largely restricted to PAs (Coad et al., 2015).

5.3. Uneven geographical representation

Geographical representation within the evidence base was uneven. Africa, Asia, and Europe collectively accounted for most studies, while MENA, South and Central America had moderate representation. In contrast, Oceania and North America were underrepresented despite their extensive PA networks and established research infrastructure. Most studies were conducted at local or regional/subregional scales, and fewer studies addressing national or multinational scales. While localised focus may limit the generalisability of findings across regions, it is useful for developing actionable, context-specific insights into protected area effectiveness. Such studies reduce the risk of drawing inaccurate conclusions from global analyses and ensure that conservation strategies are grounded in locally relevant evidence (Wyborn & Evans, 2021). On the other hand, the predominance of studies focused on a single PA reflects a tendency towards smaller-scale, site-specific assessments. While these provide detailed insights into specific locations, they may miss opportunities to evaluate broader, landscape-level impacts of PAs.

Conducting studies at broader scales could help identify cumulative impacts, connectivity issues, and broader trends that may be critical for effective conservation planning.

While all biomes are represented to some extent, the distribution is uneven, with tropical-subtropical forest and temperate-boreal forest biomes dominating the evidence base. This reflects their ecological importance as biodiversity hotspots, the global conservation priority they receive (Geldmann et al., 2013, Ge et al., 2022), and the widespread availability of forest cover data. (Adams et al., 2015). Additionally, the high concern over forest loss in many regions further drives the use of these data. However, addressing gaps in underrepresented biomes, such as alpine regions and deserts, could enhance our understanding of PA effectiveness across diverse ecological contexts, particularly for ecosystems less suited to spatial data analyses or where baseline data are more limited. This is especially relevant in the context of climate change, as ecosystems are experiencing rapid warming with significant ecological impacts (e.g. Dibari et al., 2021).

5.4. Protected area representation

The IUCN framework classifies protected areas based on management objectives, with Categories Ia, Ib, and II prioritising biodiversity conservation, often restricting human activities. Category IV focuses on habitat or species management, often requiring active interventions. In contrast, categories V and VI integrate conservation with sustainable use and have expanded significantly in recent decades (Stoudmann et al., 2023). Despite the recent growth in multiple-use PAs, our results indicate that the effectiveness for biodiversity conservation remains poorly understood compared to strictly protected areas. The predominance of studies on IUCN category II PAs may reflect their explicit conservation mandates, making effectiveness evaluations more straightforward. Multiple-use PAs, Natura 2000 sites, and UNESCO areas involve more stakeholders and varied conservation strategies, making biodiversity assessment more challenging. Their focus on balancing human use with conservation may also contribute to trade-offs that are difficult to measure (Oldekop et al., 2015). Additionally, many multiple-use PAs have only been established more recently, whereas a significant proportion of historically designated PAs fall under stricter protected categories, providing a longer track record for evaluation and assessment.

Greater attention is needed to evaluate the biodiversity effectiveness of multiple-use PAs, Natura 2000, and UNESCO sites. Future research should integrate biodiversity-focused and socioecological indicators to assess conservation trade-offs given the dual objectives of these areas. Understanding the complementary roles of different PA types can also support in advancing connectivity and landscape-level conservation approaches (Stoudmann et al., 2025). Ensuring a more comprehensive understanding of how different protected area types contribute to biodiversity conservation will help inform adaptive management strategies and strengthen global conservation efforts.

5.5. Future directions and implications for conservation

This systematic map highlights both advancements and persistent gaps in the evaluation of protected area effectiveness. While significant progress has been made in the use of spatial data, remote sensing, and control-intervention study designs, there remains a strong need for more rigorous experimental approaches, broader geographical representation, and deeper exploration of species- and genetic-level biodiversity responses. Addressing these gaps will require long-term monitoring programs, interdisciplinary collaboration, and the integration of social and ecological indicators to better understand conservation trade-offs. As global conservation targets continue to evolve, ensuring that PAs are not only expanding but also effectively conserving biodiversity will be critical. Future research should focus primarily on refining methodologies, standardising biodiversity metrics, and enhancing the evidence base to support adaptive management and policy decisions that maximise the ecological and societal benefits of protected areas. Secondly, future research should aim to integrate information on practical management factors—such as the presence or absence of management plans, the level of available resources, and the degree of stakeholder participation—in order to contribute to a more holistic understanding of protected area effectiveness.

Author contribution

Conceptualisation: FRB, NRH, SS and NS; Methodology: NRH and SS; Investigation (screening and data extraction): GA, KT and HS; Investigation (Conducting searches and retrieving articles): NRH, SS and NS; Investigation (developing and applying AI-based models): VS; Investigation (data analysis and visualization): NRH; Writing – Original Draft: NRH, SS and NS; Writing – Review & Editing: All authors; Stakeholder engagement: FRB; Project Lead: NRH and SS; Funding Acquisition: NRH, SS and NS

Conflict of interest

The authors declare no conflict of interests regarding this work.

Funding

Co-funded by the European Union under Grant Agreement N° 101052342. This work was commissioned and supervised by Biodiversa+, and produced by Liljus Ltd.

Acknowledgments

We thank Joseph Langridge, Mathieu Basille, Paul Rouveyrol, and the Biodiversa+ active partners for their insightful and constructive comments during the planning of the protocol.

References

Adams, V. M., Chauvenet, A. L. M., Stoudmann, N., Gurney, G. G., Brockington, D., & Kumpel, C. D. (2023). Multiple-use protected areas are critical to equitable and effective conservation. *One Earth*, *6*(9), pp.1173-1189.

Adams, V. M., Setterfield, S. A., Douglas, M. M., Kennard, M. J., & Ferdinands, K. (2015). Measuring benefits of protected area management: Trends across realms and research gaps for freshwater systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1681).

de Araujo Martins, C., Pays, O., Souza, F.L., Renaud, P.C., Valente-Neto, F., Silveira, M., Ochoa-Quintero, J., Provete, D.B., Santos, C.C., Melo, I. and Rodrigues, M.E. (2023). Biodiversity responses to forest cover loss: taxonomy and metrics matter. *bioRxiv*, pp.2023-05.

Borrini-Feyerabend, G., Dudley, N., Jaeger, T., Lassen, B., Pathak Broome, N., Phillips, A., & Sandwith, T. (2013). Governance of protected areas: From understanding to action. Best practice Protected Are guidelines series no. 20, Gland, Switzerland: IUCN. xvi + 124pp.

Cazalis, V., Princé, K., Mihoub, J. B., Kelly, J., Butchart, S. H., & Rodrigues, A. S. (2020). Effectiveness of protected areas in conserving tropical forest birds. *Nature Communications*, 11(1), 4461.

CBD (2022). Kunming-Montreal post-2020 Global Biodiversity Framework. CBD/COP/DEC/15/4. Convention on Biological Diversity. https://cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf

Christie, A.P., Amano, T., Martin, P.A., Petrovan, S.O., Shackelford, G.E., Simmons, B.I., Smith, R.K., Williams, D.R., Wordley, C.F.R., & Sutherland, W.J. (2021). The challenge of biased evidence in conservation. *Conservation Biology*, 35: 249-262.

Christie, A. P., Amano, T., Martin, P. A., Shackelford, G. E., Simmons, B. I., & Sutherland, W. J. (2019). Simple study designs in ecology produce inaccurate estimates of biodiversity responses. *Journal of Applied Ecology*, 56(12), 2742–2754.

Coad, L., Leverington, F., Knights, K., Geldmann, J., Eassom, A., Kapos, V., Kingston, N., de Lima, M., Zamora, C., Cuardros, I., Nolte, C., Burgess, N. D., & Hockings, M. (2015). Measuring impact of protected area management interventions: current and future use of the Global Database of Protected Area Management Effectiveness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1681), 20140281. https://doi.org/10.1098/rstb.2014.0281

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.

Collaboration for Environmental Evidence. (2022). Guidelines and Standards for Evidence synthesis in Environmental Management. Version 5.1 (AS Pullin, GK Frampton, B Livoreil & G Petrokofsky, Eds) www.environmentalevidence.org/information-for-authors. [12.01.2024]

Dawson, N., Carvalho, W. D., Bezerra, J. S., Todeschini, F., Tabarelli, M., & Mustin, K. (2021). Protected areas and the neglected contribution of Indigenous Peoples and local communities: Struggles for environmental justice in the Caatinga dry forest. *People and Nature, 3*(6), 1143–1157. https://doi.org/10.1002/pan3.10288

Dibari, C., Pulina, A., Argenti, G., Aglietti, C., Bindi, M., Moriondo, M., Mula, L., Pasqui, M., Seddaiu, G., & Roggero, P. P. (2021). Climate change impacts on the Alpine, Continental, and Mediterranean grassland systems of Italy: A review. *Italian Journal of Agronomy, 16*(3), 1843. https://doi.org/10.4081/ija.2021.1843

EUR-LEX, European Union Law. (2020). COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS EU Biodiversity Strategy for 2030 Bringing nature back into our lives, COM/2020/380 final. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0380 (Accessed: 13/11/2023).

European Commission. (n.d.). LIFE23-PRE-DE-PAME Europe on LIFE Public Database. https://webgate.ec.europa.eu/life/publicWebsite/project/LIFE23-PRE-DE-PAME-Europe-101148110/management-effectiveness-for-protected-areas-in-europe#

Ge, D., Qu, Y., Deng, T., Thuiller, W., Fišer, C., Ericson, P. G. P., Guo, B., de la Sancha, N. U., von der Heyden, S., Hou, Z., Li, J., Abramov, A., Vogler, A. P., Jønsson, K. A., & Mittermeier, R. (2022). New progress in exploring the mechanisms underlying extraordinarily high biodiversity in global hotspots and their implications for conservation. *Diversity and Distributions*, 28(7), 1375–1387.

Geldmann, J., Barnes, M., Coad, L., Craigie, I. D., Hockings, M., & Burgess, N. D. (2013). Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. *Biological Conservation*, 161, 230–238.

Geldmann, J., Manica, A., Burgess, N. D., Coad, L., & Balmford, A. (2019). A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. *Proceedings of the National Academy of Sciences (PNAS),* 116(46), 23209-23215.

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS one*, 10(9), e0138237.

Haddaway, N. R., Feierman, A., Grainger, M. J., Gray, C. T., Tanriver-Ayder, E., Dhaubanjar, S., & Westgate, M. J. (2019). EviAtlas: a tool for visualising evidence synthesis databases. *Environmental Evidence*, 8, 1-10.

Haddaway, N. R., Grainger, M. J., & Gray, C. T. (2022). Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Research Synthesis Methods*, 13(4), 533-545.

Haddaway, N.R., Stoudmann, N. and Savilaakso, S., 2024. How is the effectiveness of terrestrial protected areas to conserve biodiversity measured? A systematic map. https://doi.org/10.32942/X2762M

Hansen, M.C., Stehman, S.V. and Potapov, P.V. (2010). Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences*, *107*(19), 8650-8655.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R. and Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850-853.

Hockings, M., Stolton, S., Leverington, F., Dudley, N. and Courrau, J. (2006). Evaluating Effectiveness: A framework for assessing management effectiveness of protected areas. 2nd edition. IUCN, Gland, Switzerland and Cambridge, UK. xiv + 105 pp.

Hofmann, D. D., Behr, D. M., McNutt, J. W., Ozgul, A., & Cozzi, G. (2021). Bound within boundaries: Do protected areas cover movement corridors of their most mobile, protected species? *Journal of Applied Ecology*, 58(6), 1133-1144.

Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (n.d.). Policy support tool: Protected Areas Management Effectiveness (PAME) evaluations. https://www.ipbes.net/fr/node/15618

James, K.L., Randall, N.P. and Haddaway, N.R., 2016. A methodology for systematic mapping in environmental sciences. *Environmental Evidence*, 5, pp.1-13.

Li, G., Fang, C., Watson, J.E.M., Sun, S., Qi, W., Wang, Z., and Liu, J., 2024. Mixed effectiveness of global protected areas in resisting habitat loss. Nature Communications, 15, 8389. https://doi.org/10.1038/s41467-024-52693-9

Mascia, M. B., Pailler, S., Thieme, M. L., Rowe, A., Bottrill, M. C., Danielsen, F., Geldmann, J., Naidoo, R., Pullin, A. S., & Burgess, N. D. (2014). Commonalities and complementarities among approaches to conservation monitoring and evaluation. *Biological Conservation, 169, 258-267.*

Maxwell, S. L., Cazalis, V., Dudley, N., Hoffmann, M., Rodrigues, A. S. L., Stolton, S., Visconti, P., Woodley, S., Kingston, N., Lewis, E., Maron, M., Strassburg, B. B. N., Wenger, A., Jonas, H. D., Venter, O., & Watson, J. E. M. (2020). Area-based conservation in the twenty-first century. *Nature*, 586(7828), 217–227.

Oldekop, J.A., Holmes, G., Harris, W.E. and Evans, K.L. (2016), A global assessment of the social and conservation outcomes of protected areas. *Conservation Biology*, 30: 133-141.

Potapov, P., Yaroshenko, A., Turubanova, S., Dubinin, M., Laestadius, L., Thies, C., Aksenov, D., Egorov, A., Yesipova, Y., Glushkov, I. and Karpachevskiy, M. (2008). Mapping the world's intact forest landscapes by remote sensing. *Ecology and Society*, *13*(2).

Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R.D. and Honrado, J.P. (2017). Global biodiversity monitoring: from data sources to essential biodiversity variables. *Biological Conservation*, *213*, 256-263.

Rodrigues, A. S. L., & Cazalis, V. (2020). The multifaceted challenge of evaluating protected area effectiveness. *Nature Communications*, 11(1), 5147.

Rodríguez-Rodríguez, D., & Martínez-Vega, J. (2022). *Effectiveness of protected areas in conserving biodiversity: a worldwide review.* Springer Nature. ISBN: 303094297X. 128 pp.

Smokorowski K. E., & Randall, R. G. (2017). Cautions on using the Before-After-Control-Impact design in environmental effects monitoring programs. *FACETS*, 2, 212-232.

Stewart, F. E., Darlington, S., Volpe, J. P., McAdie, M., & Fisher, J. T. (2019). Corridors best facilitate functional connectivity across a protected area network. *Scientific Reports*, 9(1), 10852.

Stoudmann, N., Byrne, J., & Adams, V. (2025). Effect of reserve protection level and governance on tree cover loss and gain. *Conservation Biology*, e14449.

Stoudmann, N., Savilaakso, S., Waeber, P. O., Wilmé, L., Garcia, C., Byrne, J., & Adams, V. M. (2023). Overview of evidence on mechanisms affecting the outcomes of terrestrial multiple-use protected areas. *One Earth*, *6*(5), 492–504.

Visconti, P., Butchart, S. H. M., Brooks, T. M., Langhammer, P. F., Marnewick, D., Vergara, S., Yanosky, A., & Watson, J. E. M. (2019). Protected area targets post-2020: Outcome-based targets are needed to achieve biodiversity goals. *Science*, 364(6437), 239–241.

Watson, J. E., Dudley, N., Segan, D. B., & Hockings, M. (2014). The performance and potential of protected areas. *Nature 515*, 67-73.

Wyborn, C., & Evans, M.C. (2021). Conservation needs to break free from global priority mapping. *Nature Ecology & Evolution 5*, 1322-1324.

Annex 1. A list of articles to test comprehensiveness of the search string

1. Andam, K. S., Ferraro, P. J., Pfaff, A., Arturo Sanchez-Azofeifa, G., & Robalino, J. A. (2008). *Measuring the effectiveness of protected area networks in reducing deforestation*. www.pnas.orgcgidoi10.1073pnas.0800437105
2. Carrillo, E., Wong, G., & Cuarón, A. D. (2000). Monitoring Mammal Populations in Costa Rican Protected Areas under Different Hunting Restrictions. In *Conservation Biology* (Vol. 14, Issue 6). https://www.jstor.org/stable/2641510
3. Gardner, C. J., Jasper, L. D., Eonintsoa, C., Duchene, J. J., & Davies, Z. G. (2016). The impact of natural resource use on bird and reptile communities within multiple-use protected areas: evidence from sub-arid Southern Madagascar. *Biodiversity and Conservation*, *25*(9), 1773–1793. https://doi.org/10.1007/s10531-016-1160-4
4. Graham, V., Geldmann, J., Adams, V. M., Negret, P. J., Sinovas, P., & Chang, H. C. (2021). Southeast Asian protected areas are effective in conserving forest cover and forest carbon stocks compared to unprotected areas. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-03188-w
5. Kallimanis, A. S., Touloumis, K., Tzanopoulos, J., Mazaris, A. D., Apostolopoulou, E., Stefanidou, S., Scott, A. v., Potts, S. G., & Pantis, J. D. (2015). Vegetation coverage change in the EU: patterns inside and outside Natura 2000 protected areas. *Biodiversity and Conservation*, *24*(3), 579–591. https://doi.org/10.1007/s10531-014-0837-9
6. Knorn, J., Kuemmerle, T., Radeloff, V. C., Keeton, W. S., Gancz, V., Biriş, I.-A., Svoboda, M., Griffiths, P., Hagatis, A., & Hostert, P. (2013). Continued loss of temperate old-growth forests in the Romanian Carpathians despite an increasing protected area network. *Biodiversity Governance in Central and Eastern Europe*, *40*(2), 182–193. https://doi.org/10.2307/26319125
7. Pfeifer, M., Burgess, N. D., Swetnam, R. D., Platts, P. J., Willcock, S., & Marchant, R. (2012). Protected areas: Mixed success in conserving East Africa's evergreen forests. *PLoS ONE*, *7*(6). https://doi.org/10.1371/journal.pone.0039337
8. Terraube, J., Gardiner, R., Hohwieler, K., Frère, C. H., & Cristescu, R. H. (2023). Protected area coverage has a positive effect on koala occurrence in Eastern Australia. *Biodiversity and Conservation*, *32*(7), 2495–2511. https://doi.org/10.1007/s10531-023-02615-w
9. Wauchope, H. S., Jones, J. P. G., Geldmann, J., Simmons, B. I., Amano, T., Blanco, D. E., Fuller, R. A., Johnston, A., Langendoen, T., Mundkur, T., Nagy, S., & Sutherland, W. J. (2022). Protected areas have a mixed impact on waterbirds, but management helps. *Nature*, *605*(7908), 103–107. https://doi.org/10.1038/s41586-022-04617-0

Annex 2. Definitions of meta-data values

1. Sampling methodology

Transects	A straight line that cuts through a natural landscape so that standardised observations and measurements can be made
Sampling plots	A clearly defined area of land where sampling units for one or more groups of taxa are located and measured
Point counts	A tally of observations detected by sight and/or sound by a single observer located at a fixed position during a specified period of time
Stratified sampling	Sampling from divisions across a habitat, which appear as different zones
Mark-recapture	Used to estimate the size of a population where it is impractical to count every individual. A small number of individuals is captured, marked, released and then recaptured systematically
Random sampling	Samples selected randomly from the whole area
Camera traps	Cameras are positioned in a static location and movement triggers recording of a series of images of identifiable species/individuals
Geographical Information System	Including satellite and aerial imagery
Systematic sampling	Collecting data in an ordered or regular way, eg every 5 metres or every fifth tree

2. Survey method (outcome measurement)

pit-traps	A trapping pit for small animals, such as insects, amphibians and reptiles
basket traps	Traps for freshwater organisms that facilitate entry but prohibit exit (made as baskets)
kick sampling	A kick net is held against the riverbed with the water flowing into it, whilst upstream of the net, the surveyor kicks the riverbed to disturb and dislodge any invertebrates
satellite measurement	Satellite or aerial imagery is recorded and used to quantify electromagnetic radiation (visible/non-visible light)
sweep netting	Sturdy nets, often with a canvas bag, that are used to collect insects and other invertebrates from long grass
hand searching	Manually turning over leaves or stones to search for organisms
direct observation	Measuring something by direct line of sight
indirect observation	Measuring something by using indications of presence (e.g. prints or scat)
direct listening	Measuring something by listening for calls/noise of presence
cast net	A fishing net that is thrown out and immediately drawn in again, rather than being set up and left
gillnet	A static fishing net which is hung vertically so that fish get trapped in it by their gills
active fishing	Fishing gear that is dragged through the water by human, animal or engine power
live trap	A terrestrial box used to trap mobile organisms
questionnaire	An interview survey given to humans

3. Study type

Experimental	Something in the environment is directly manipulated by researchers and then measured
Observational	A system is observed without intentionally being affected in any way

4. Experimental design

BACI	Before-After-Comparator-Intervention (i.e. at least one control sire measured JUST before and after the intervention happened, i.e. not multiple measurements over time after intervention)
BA	Before-After - i.e. no more than 2 measurements, one before and one after the intervention was put in place
CI	Comparator-Intervention - i.e. a control site and an intervention site measured ONLY after the intervention happened (no more than 2 measurements over time, otherwise it is "Time-series")
Time-series	Time-series - i.e. 3 or more measurements of the intervention taken over time (NO CONTROL sites)
CI + time-series	Comparator-Intervention and time-series together - i.e. a control and intervention site measured more than 3 times ONLY AFTER the intervention occurred (NO BASELINE)
BACI + time-series	Before-After-Comparator-Intervention and time series - i.e. baseline measurements before the intervention at a control and an intervention site, measured 3 or more times in total (not just twice, before and after)