Who reads the Earth's archives? Geography, Income, and the Uneven Evidence Base for a Shared Planet

Lingli Fan [a], Yiming Lei [a], Yufei Chen [a,b], Qian Huang [d*], Hui Jia [e], Shan Jiang [f], Jingchun Liu [a], Haoliang Lu [a], Chongling Yan [a], Hualong Hong [a,b,c*]

Affiliations

a. Key Laboratory of Ministry of Education for Coastal and Wetland Ecosystems, Xiamen University, Xiamen 361102, China

b. Coastal and Ocean Management Institute (COMI), Xiamen University, Xiamen 361102, China

c. National Observation and Research Station for the Taiwan Strait Marine Ecosystem (Xiamen University), Zhangzhou 363305, China

d. Institute of Geosciences, University of Mainz, Johann-Joachim-Becher-Weg 21, Mainz 55128, Germany

e. School of the Environment and Safety Engineering, Jiangsu University, Zhenjiang 212013, China

f. State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China

Author email addresses and ORCID

Lingli Fan (a) — 33120251152187@stu.xmu.edu.cn
Yiming Lei (a) — 33120241152277@stu.xmu.edu.cn
Yufei Chen (a,b) — 33120241152351@stu.xmu.edu.cn
Qian Huang (d) * — qhuang01@uni-mainz.de
Hui Jia (e) — 1000005403@ujs.edu.cn
Shan Jiang (f) — sjiang@sklec.ecnu.edu.cn
Jingchun Liu (a) — liujingchun@xmu.edu.cn
Haoliang Lu (a) — luhl@xmu.edu.cn, 0000-0003-0720-3137
Chongling Yan (a) — ycl@xmu.edu.cn
Hualong Hong (a,b,c) * — honghl@xmu.edu.cn, 0000-0003-4284-8096

Abstract

Global environmental assessments depend on natural-archive evidence, yet where this evidence is produced and how its geography maps onto the Sustainable Development Goals (SDGs) remains unclear. Here we map the sampling and publication geography of 33,145 studies of past environments, tracing each to the SDG it addresses. The geography is highly unequal. A group of 26 countries is home to 57.5% of world population but accounts for 81% of sampling and 89% of publication credit, with per capita evidence output about 28 times higher and all eleven goals drawing 82–90% of their evidence from them. Income acts as a threshold rather than a gradient, with output flat below about 41,000 PPP dollars per capita and rising steeply above. Sampling and publication are concentrated in the same countries, so sample accessibility is unlikely to be the main constraint; the deficit is more consistent with limited analytical capacity. The evidence base for the SDGs therefore reflects a minority's research geography, not the distribution of the processes measured. Our findings provide a baseline for measuring commitments to widen participation and identify analytical capacity, not sample access, as the binding constraint on whose evidence is used in global assessments.

Keywords: Bibliometrics; Palaeoclimatology; Research inequality; Geographic bias; Open science

1 Introduction

Research inequality is a structural property of global science, not a residual of national differences in talent. Economic scale and cumulative advantage reinforce it. National output rises with income[1,2], and funding, instruments and trained researchers accumulate where they are already concentrated, so that past success predicts future success independently of the quality of the work[3,4]. Colonial histories and present evaluation regimes organise its geography, concentrating interpretive capacity in the countries that also wrote the rules of what counts as evidence[5–8]. If capacity is the binding

constraint, the countries that can analyse should dominate publication. Whether they also dominate sampling is the empirical question that separates extraction from exclusion. Income should then act as a threshold rather than a gradient, and the evidence base for global goals should inherit the geography of capacity rather than the geography of the processes being measured. Natural archives are the decisive case in which to test this.

Natural archives are the evidence system through which past environmental change enters global assessments. They record change beyond instrumental records[9], supply the baseline against which contemporary change is judged[10], and underpin model evaluation[11,12], estimates of industrial-era warming[13,14] and assessments of cryosphere response[15]. Standardised dating, age–depth modelling and community data infrastructure have made this literature comparable across sites[16–20], and a broader analytical repertoire has extended it to speleothems and loess[21,22]. That comparability is what makes a global analysis of its geography possible. Because assessments inherit this literature, the organisation of archive research is a policy variable rather than a curiosity of disciplinary history.

The literature has measured the pieces of this geography separately, but not their coincidence. Bibliometric analysis describes publication dynamics, national rankings and thematic structure from complete corpora[23–25], typically through co-citation and keyword mapping[26], but it rarely asks who can analyse the material. Research on parachute and helicopter science documents a different pattern at site level. Material is collected in one country and analysed in another, and publication counts overstate the capacity of sampling-rich countries[27–30]. Neither literature tests whether sampling and publication are coupled at country level, whether income acts as a threshold, or whether the resulting concentration reaches the evidence attached to global goals. The present study supplies that test.

Here we reconstruct the sampling and publication geography of 33,145 studies of past environments from the Web of Science Core Collection and OpenAlex, and trace each record to the Sustainable Development Goal it informs. We ask three questions. The first concerns how unequal the field is

across countries and archive types, and whether income acts as a threshold rather than a gradient. The second concerns how the hierarchy of archive carriers, proxy indicators, scientific questions and societal goals organises that inequality. The third, and most consequential, concerns whether the concentration reaches the evidence on which global goals depend. Answering them requires a measure of sampling independent of publication counts, a country classification built on both dimensions, and a record-level network that preserves the multi-label character of the data.

2 Results

2.1 Production and sampling are jointly concentrated

Production and sampling are tightly coupled across countries and both are extremely concentrated (Figure 1a-c). Cross-classifying 195 countries gives 26 in the high-high class and 39 countries with zero publications and zero sampling, and two of the nine cross-classes are empty, so no country publishes heavily without sampling or samples heavily without publishing. The two dimensions correlate at Spearman rho = 0.884 and Pearson r = 0.938, and the high-high class, 13.33% of countries, accounts for 89.11% of fractional publications and 81.29% of sampling-location articles. The Gini coefficient is 0.910 for publications and 0.847 for sampling. The concentration has a continental expression, with Europe leading on publications (45.1% of fractional credit) and Asia on sampling (34.3%), and the ten largest producers run from the United States to the Russian Federation.

The signed wedge between publication and sampling shares is small everywhere and robust to how multi-label records are treated (Figure 1d, e; Table S3). It is indistinguishable from zero for all archive sub-classes tested, and assigning every record to its first sub-class instead leaves the two sediment Ginis unchanged at 0.8748 and 0.8559. The largest negative median is polar ice cores, at -0.0003, but no sub-class differs significantly from zero. Across the sub-classes the size of the wedge

tracks how widely an archive is sampled rather than how many records it carries or how expensive the material is to obtain.

The asymmetry is not uniform across archive types (Figure 1d, e; Tables S3, S4). Marine and lake sediments are the closest available natural experiment, being alike in material and indicator repertoire and differing mainly in whether sampling requires a research vessel or a drilling programme. Their sampling concentration differs by +0.0189 Gini points, marine 0.8748 against lake 0.8559 with a 95% bootstrap interval of 0.0133 to 0.0237, so facility dependence does contribute to inequality. Polar ice cores are more concentrated still, 0.9742 against 0.9052 for the distributed tree-ring network. Their binding constraint is not cost but sovereignty, of the 2,824 polar ice-core records, 1,388 were sampled in Antarctica and 642 in Greenland, and 2,265 carry at least one location outside the 195-country roster, because Antarctica has no sovereign state and Greenland is not one. The archive that is hardest to reach is also the one whose sampling lies furthest outside the system of sovereign states.

The layers differ sharply in how narrowly they are held (Figures S2, S3). Archive classes are strongly unequal in size, with sediment carrying 70.85% of all labelled records against 0.17% for other natural carriers, and no record matches two or more top-level classes. Indicators concentrate more narrowly than questions, biological material leading at 39.72% against 33.86% for chemical composition, the two classes that together account for most of the field's proxy repertoire[31,32], whereas cognitive goals are stable over time and led by reconstruction and inversion at 33.49%.

2.2 Income, scaling and the elasticity step

Income separates the corpus more sharply than geography (Figure 2a). High-income countries, 64 of 195, produce 70.52% of fractional publications against 0.09% for the 25 low-income countries, and the gradient is monotonic across the four World Bank groups. Income must be measured in constant

purchasing power, because the same regression returns an exponent of 1.009 on current US dollars and 1.351 on constant 2021 PPP international dollars, a difference that changes the shape of the relationship rather than its scale.

The residuals of the country-level fit are not symmetric, and the asymmetry is concentrated at entry rather than spread across the low end (Figure 2b-d). Grouping by fractional output, the median residual is -0.383 log1p units for countries with no publications, -0.638 for the one-to-five group, +0.921 for the 6-20 group and +0.258 for the 21-100 group, and indistinguishable from zero only above 100 publications (-0.040). The intervals for the one-to-five, 6-20 and 21-100 groups exclude zero. The 6-20 group is therefore above the fitted line, so the deficit is specific to countries entering the literature rather than a general low-end shortfall. The relation is also not a single power law, sampling is a zero-inflated response, with 85 of 195 countries recording none and 42 publishing nothing, and a Tweedie generalised additive model needs 3.52 effective degrees of freedom where a straight line would use one.

A single exponent is nevertheless a poor description, because elasticity is a property of countries rather than of the field (Tables S2, S5). Country-specific elasticities span more than an order of magnitude and 42 of the 56 countries with enough observations lie above the pooled value. They are not randomly distributed, since a model in which the elasticity is flat below an income threshold and steps upward above it fits far better than a single trend (AIC 289.6 against 307.7). The step, about 3.96 elasticities at 41,458 PPP international dollars per head, is the whole of the signal (Figure 2e), because once it is in the model the residual slope is indistinguishable from zero while the step interval excludes it.

2.3 How inequality propagates into archives, questions and goals

The network is built from the 29,379 records whose first-author country has a complete GDP per capita series, which is 88.6% of the corpus and 95.4% of fractional publications (Figure 3). Countries are split at the breakpoint of 41,458 PPP international dollars identified in section 2.2, so 33 countries lie above it and 62 below. The two groups are unequal in output. Countries above the breakpoint carry 20,811 records, 70.8% of the flow, against 8,568 and 29.2% for those below. Because each record contributes a total weight of one and is split evenly across the labels it carries, every layer sums to the same 29,379 records, so the flow axis reads directly as records. Archive composition is led by sediment at 71.1%, then biological growth at 13.7% and ice core at 8.3%. Four goals absorb 95.9% of the flow, SDG 13 (Climate Action) taking 60.7%, SDG 15 (Life on Land) 13.0%, SDG 14 (Life below Water) 12.3% and SDG 6 (Clean Water and Sanitation) 9.9%.

The layers are unequally concentrated, and unequally informative. Indicators are led by biological material at 39.5% and concentrate more narrowly than questions, whose leading category carries 33.3%, and the same ordering holds by the Herfindahl index, 0.211 for questions against 0.307 for indicators, and by normalised entropy, 0.806 against 0.728. The community converges more on what to measure than on what to ask. The layer pairs differ even more sharply in how far they constrain one another. Measurement follows the material closely, archive sub-class and indicator class sharing 9.3% of the entropy of the latter, by far the strongest association in the network. Question type is nearly independent of both archive sub-class and indicator class, at 1.5% or less, which is the quantitative form of the claim that what is asked does not follow from what is measured.

The concentration is mainly a matter of capacity rather than of agenda (Figure 3b). For each layer the penetration index gives log2 of the ratio between the share a node holds among records above the breakpoint and the share it holds below it, and the total variation distance summarises how far the two distributions diverge. The clearest over-representation is in the means of research rather than its ends, ice core at +1.64 and method development at +1.47. The goal layer has the smallest total

variation distance of the four, 0.022 against 0.103 to 0.120 elsewhere, but that aggregate conceals a real split, because SDG 13 carries 61% of the flow and differs between the two groups by only 1.14 percentage points with p = 0.071. The smaller goals do diverge. SDG 11 Sustainable Cities is over-represented 1.8-fold, with 203 additional records and p = 3.6e-08, and SDG 14 Life below Water is under-represented by 256 records with p = 0.004. High-capacity countries lean towards cities, industry and climate and low-capacity countries towards hunger, health and water, while both place more than 95% of their output on SDGs 13, 15, 14 and 6.

3 Discussion

The field's inequality is severe, and it is a capacity gap rather than a sampling gap. Twenty-six countries supply 89.11% of its publications and 81.29% of its sampling, and the countries outside that group are marginal in both, so what separates the field is not who collects samples but who can analyse them. That distinction decides where reform should aim, and it frames the three questions this discussion takes up, namely what produces the inequality, what it costs the evidence base on which global policy depends, and what would close it. Income separates the corpus more sharply than geography, and the elasticity of output to income is a step rather than a gradient, indistinguishable from zero below about 41,458 PPP international dollars per head and jumping roughly fourfold above it. Because archive classes are separate communities with separate geographies, and because the knowledge network funnels 95.9% of its flow into four Sustainable Development Goals, concentrating capability narrows the questions the field is able to answer.

3.1 A capacity gap, not a sampling gap

The coupling between sampling and publication is a diagnostic, not the finding. The parachute-science literature documents extraction. Material is collected in one country and analysed in another, and field sites are compared with author affiliations27–29. At country level that mismatch disappears,

because the countries that publish are the countries whose territory is sampled. What remains is a different inequality, between the small group that does both and the large group that is marginal in both. Framing the problem as decoupling would point policy towards sample-sharing agreements; framing it as exclusion points towards building analytical capacity where sampling already happens. The stratified analysis shows where that exclusion is strongest.

Income separates the field more sharply than geography, and only constant purchasing power reveals the shape of the relation rather than its scale. Elasticity is a property of countries rather than of the field. 42 of 56 well-observed countries lie above the pooled value, and the step at 41,458 PPP international dollars per head is the whole of the signal. The step survives normalising output by population, which accounts for about 8% of the raw-total step (Tables S2, S5). What matters is not the exponent but that countries below the threshold convert resources into publications at the same, near-zero rate, so no amount of income below it predicts who publishes.

The shape of the inequality identifies the mechanism more precisely than the correlation does. Comparing each class's share of publications with its share of sampling gives a signed wedge. The high-high countries hold 89.11% of publications but 81.29% of sampling, a surplus of 7.8 points, while the 39 medium-medium countries run a deficit of 7.2 points. Both wedges are small relative to the concentration itself. The Gini coefficient is 0.910 for publications and 0.847 for sampling. Some of the core's output therefore rests on archives outside its territory, but the larger fact is that 39 countries are absent from both dimensions, and many more are marginal in both. Four mechanisms can produce that pattern. Cumulative advantage concentrates funding, instruments and trained researchers where they already are[3,4], and unequal access to methods raises the return to existing capacity[1,2]. A colonial division of labour routes material and credit towards the core where infrastructure is costly[27,30,33], and training and evaluation reproduce the questions the core is equipped to ask[7,34]. Credit follows the same gradient, because citation rates rise with national wealth

and governance, not only with the importance of the research[35]. These mechanisms are not mutually exclusive, and this design cannot weight them; what they share is a prediction of concentrated interpretive capacity, not of scarce samples.

One alternative explanation has to be ruled out, namely that the coupling is a bookkeeping identity, if every paper were written by an author from the country where its sample was taken, the two counts would move together by construction. Three observations argue against this. A purely mechanical relation would leave every residual at zero, whereas the one-to-five group sits below the fitted line and the 6-20 group above it, with intervals that exclude zero; it would give a correlation of one, whereas the observed Pearson r of 0.938 leaves a measurable wedge; and it would not make the sign of the wedge vary systematically with class, yet the wedge is positive at the top and negative in the middle and lower tiers. What the design cannot exclude is that a single common cause, the size of a national research system, drives both dimensions, and income is only a partial proxy for it.

If capacity is the common cause, it need not act equally on the means and ends of research. The penetration analysis separates the two. High-capacity countries are over-represented in the means of research—ice cores (+1.64) and method development (+1.47)—more than in its stated ends. The goal layer has the smallest total variation distance (0.022), and SDG 13 differs between the two groups by only 1.14 percentage points. Divergence appears in smaller goals, with SDG 11 over-represented 1.8-fold and SDG 14 under-represented by 256 records. This is consistent with partial control over how the field measures rather than over what it asks, and with accounts in which research priorities are mediated by funders, disciplines and evaluation regimes rather than set by societal demand alone[34]. It is not a direct measurement of agenda control, and the country-level design cannot establish who within a country sets the question.

3.2 The evidence base for global goals is not globally produced

Concentration would be a matter of scholarly geography were it not for what the network delivers. Because every record resolves to a Sustainable Development Goal, the network measures not only who publishes but who supplies the evidence attached to each goal, and that distribution is narrower than the publication distribution it derives from. All eleven goals with non-zero flow draw between 81.9% and 90.1% of their evidence from the twenty-six high-high countries, while lower-middle-income countries contribute 0.00% to every goal and low-income countries between 0.00% and 0.12% (Figure S6). The gap is one of intensity as well as totals, since those countries hold 57.5% of world population and per head produce the evidence reaching these goals roughly twenty-eight times more intensively than the poorest classes combined.

This matters because the assessments that inform the Sustainable Development Goals are built from the same literature this network describes. Where that literature comes from a narrow set of countries, its spatial and ecological coverage inherits the research geography of those countries rather than following the distribution of the processes being measured, a limitation the Intergovernmental Panel on Climate Change (IPCC) has itself recognised in its calls for geographic balance[11]. The point concerns the evidence base and not the quality of individual studies. Representation is also an epistemic problem. Authorship shapes which evidence is treated as relevant, and assessments with narrow author pools can reproduce narrow content[36]. Language compounds the bias, because non-English evidence is systematically underused even where it is directly relevant[37,38], and sampling itself is uneven[39–41]. It also decides the remedy, because if the deficit were one of sampling access the answer would be agreements that share material, whereas a deficit of analytical capacity has to be met by building that capacity where the archives and the affected populations are.

That diagnosis locates the gap in analytical capacity, and the international apparatus is not aimed at it. The governance gap is structural, because instruments written for genetic resources assume that value and benefit flow from access to a physical sample, whereas natural archives yield their value

chiefly through analysis. The CBD and its Nagoya Protocol[42,43], the Kunming-Montreal Global Biodiversity Framework[44] and the decisions on digital sequence information[45–47] regulate the movement of material or its digital representation, and the agreement on marine biodiversity beyond national jurisdiction[48] follows the same logic. Extending an access-and-benefit-sharing architecture to natural archives therefore cannot by itself close the gap.

The generic instruments are also asymmetric between reading and writing. The UNESCO Recommendation on Open Science names equity and capacity building among its commitments[49], against a baseline in which four out of five countries invest less than one per cent of GDP in research[50]. Research4Life and the diamond open-access sector address access to the literature[51,52], the Systematic Observations Financing Facility (SOFF) pays for observing capacity in the least developed countries[53] and the UN Decade of Ocean Science runs a capacity workstream[54]. Open-science reforms can widen rather than narrow the gap when they shift costs onto authors and institutions without the infrastructure to absorb them[55]. Editorial calls for geoscience to globalise, and evidence that authors from low- and lower-middle-income countries remain under-represented even in journals committed to equitable development, describe the same asymmetry[56,57]. Nothing among them targets the conversion of sampling capacity into publication output, and the countries with one to five publications sit exactly on that boundary. Partnership evidence points to in-country leadership and long-term institutional support as the missing ingredients[33,58].

3.3 What would close the gap

The field nevertheless holds levers that are specific to it, and they are more concrete than the generic instruments. Training is the first, because the elasticity step is a capacity threshold rather than a funding one, and the countries just below it already have the archives and the field programmes. Its concrete form is the twinned doctoral and postdoctoral programme, in which a student is registered

in a laboratory in a high-high country and a laboratory in a low-publication country and divides their time between them, and the constraint is not the willingness of either partner but the absence of a funding line that treats analysis, rather than sampling, as the object of capacity building. Evidence from North–South partnerships supports this design. What sustains capacity is long-term institutional partnership, in-country leadership and shared infrastructure, rather than short courses alone[59,60]. Regional synthesis networks are the second, since they let countries with archives but thin analytical capacity contribute to questions that would otherwise be answered elsewhere; the community databases already built for this purpose, Neotoma and PANGAEA among them, show that the infrastructure exists[61–63]. Governed access is the third, and it exists but is not equitable, as ocean drilling protects new cores with a one-year post-cruise moratorium while land-based archives carry no equivalent[64]. The deep-time initiatives that have tried to build the missing capacity at scale remain the exception[65,66], and the most recent global assessment reaches the same conclusion[67].

Shared infrastructure is the fourth lever, and the largest attempt to build it is Chinese-led. Across this corpus the share of publications produced by the twenty-six high-high countries falls only from 93.2% in the first period to 88.9% in the last, so the question is where the chain is weakest rather than which link to abandon. Sample access, repository and curation systems, and analytical laboratories are complementary rather than substitutable, and strengthening one without the others leaves the chain no stronger; what the evidence here identifies is that the analytical link is the binding one, not that the others do not matter.

The community's own governance is where an equity norm would have to live, and it is half built. Past Global Changes requires that the data behind any publication it acknowledges be publicly available and encourages inclusive authorship of large syntheses[68], but its code of conduct governs meetings rather than fieldwork or sampling, In June 2025 Past Global Changes discontinued the mobility fellowships aimed at early-career scientists in low- and middle-income countries. The most

transferable instrument is a journal policy. Climate of the Past requires authors to address power inequalities at the proposal, data-access, fieldwork, publication and legacy stages, including in-country leadership of publications, and grounds the requirement in the geoscience union's 2023 statement recognising helicopter research as scientific neocolonialism[69]. A journal in this field could adopt the same architecture, with sample-and-data return, in-country collaboration and authorship norms, placing the step this paper isolates inside a reviewer-facing obligation. Access to an archive on national territory is granted by permit, and a permit condition can require a named local collaborator, deposit of a reference collection in a national repository, return of derived data or registration of the samples, as biodiversity research has begun to do[7,30,70], and where sampling overlaps Indigenous territories the CARE principles set the standard for the data side[71]. A permit condition therefore acts where the asymmetry lies.

If this interpretation is correct, the gap should close in specific and measurable ways. The elasticity step should flatten as analytical capacity spreads; the residual of the one-to-five group should move towards zero and the positive residual of the 6-20 group should narrow; the number of countries with at least one fractional publication, 153 of 195 here, should rise; the share of flow reaching prediction and simulation, at present 0.33%, should increase; and the wedge between publication and sampling shares at the top of the distribution should narrow. Each is computable from the next release of the corpus, and we report them here as a baseline rather than a forecast.

Three options follow for intergovernmental bodies already active in this space, and all work within existing mandates. The first is to make analytical capacity an explicit criterion in the capacity-building workstreams they fund, so that a proposal which builds a laboratory or a twinned programme competes on the same footing as one that buys a ship or a drill. The second is to attach an authorship condition to sample access, requiring that work derived from material collected under a national permit includes authors from the country of origin, which converts an access regime into a

capacity instrument without changing its legal architecture. The third is to extend the curation duties that data repositories already carry to the laboratory partnerships that generate measurements, so that the instruments and the training that produced a dataset are recorded alongside it and can be credited and repeated.

3.4 Limitations and outlook

Four limitations bound these results. The corpus is a union of the Web of Science Core Collection and OpenAlex, because WoS alone under-covers non-English and non-indexed work; the union is more complete but remains biased towards English-language and indexed literature; a further 16,044 records retrieved by the same queries were excluded as out of scope by a model judgement, so the analysed corpus is a screened rather than an exhaustive set (Table S7). Field labels are model-derived, and 2.54% of records could not be assigned to an archive class while 19.23% of sampling-location fragments were not resolvable to a country, so archive shares describe the classified subset. Country attribution rests on authorship locations under fractional counting and on stated sampling locations, and neither captures who owns a sample or its data.

The definition of sampling is itself a choice with consequences. It is taken throughout as the geography of the archives a paper studies rather than the fieldwork its authors performed, which is why 1,293 reviews and syntheses enter the corpus. Excluding them moves the sediment and ice-core contrasts by 0.0001 and 0.0005 Gini and sampling concentration by 0.0006, so the choice does not drive the conclusions (Table S8); the stratified analysis, which requires an archive sub-class and a resolvable sampling country, retains 32,302 of the 33,145 records, and the unclassified remainder is not randomly distributed over time.

The interpretation the paper rests on is an inference rather than a measurement. Research inequality is also multidimensional. Gender, language, institution and career stage cut across the national

355 geography measured here, and a country-level design cannot observe within-country inequality or
356 attribute it to particular actors[35,37,38]. That the pattern is a capacity gap rather than a sampling gap
357 follows from two measured quantities, publication and sampling, and the network, funding and
358 training variables that would measure analytical capacity directly are not observed here. The
359 Sustainable Development Goal layer encodes goals named in abstracts rather than verified
360 contributions, and the sampling year is the publication year, a proxy for sampling activity rather than
361 a record of it.

362 Three lines of work follow. First, the coupling result should be tested at site level, where the
363 parachute-science literature operates, to establish whether the country-level coupling survives finer
364 resolution. Second, the relation between income and output should be examined with instrumentation
365 and funding variables rather than income alone, since the contrast between marine and lake
366 sediments shows that the threshold is not simply the cost of field equipment, and they need to
367 separate the capacity to acquire material from the capacity to analyse it. Third, the network should be
368 re-estimated as the corpus grows, so that the topology reported here can serve as a baseline against
369 which capacity-building interventions are measured.

370 **3.5 Conclusions**

371 From 33,145 records assembled from the Web of Science Core Collection and OpenAlex, 31,964 of
372 them published between 1980 and 2025, this study reconstructs the global structure of
373 paleoenvironmental natural-archive research. The field's evidence is produced by a minority.
374 Twenty-six countries account for 89.11% of fractional publications and 81.29% of sampling-location
375 articles, the two extreme mixed classes are empty, and income separates the corpus more sharply
376 than geography alone. Sampling and publication are concentrated in the same countries, which is
377 why the deficit is one of analytical capacity and exclusion rather than of sample extraction. Output

grew from 26 articles in 1980 to a plateau averaging about 1857.8 per year, following the saturating form long expected of quantitative studies of science[72], passing through four periods of 2,489, 5,685, 8,471 and 15,319 articles respectively; a further 1,181 records carry publication years of 2026 or 2027 and fall outside the analysis window, which is why the four periods sum to 31,964 rather than to the full corpus. Archive physics constrains what can be measured but not what is asked, and societal goals are reached through a narrow set of question types rather than a single route, with reconstruction and inversion carrying 33.49% of the flow and mechanism and drivers 24.26%.

Income-output elasticities are not a property of the field, since they differ by more than an order of magnitude between countries and follow a step rather than a gradient, below an income threshold of 41,458 PPP international dollars per head they are indistinguishable from zero, and above it they jump by roughly four. The consequence is not merely uneven credit but an uneven evidence base. Every Sustainable Development Goal draws between 81.9% and 90.1% of its evidence from those same 26 countries, and per head of population the evidence reaching them is produced roughly twenty-eight times more intensively in the high-high class than in the lower-middle, low-income and low-low classes combined (Figure S6). The record that global environmental assessments rest on is therefore shaped by the research geography of a minority of the world's population rather than by the distribution of the processes being measured. The instruments that exist for closing the gap, open-science policy and access and benefit-sharing, mostly regulate access to samples and data, whereas the binding constraint identified here is analytical capacity in the countries that host the archives. Building that capacity is what would let the evidence base for a shared global future represent the planet rather than the part of it equipped to measure it; future work should test whether such investment changes the topology of this network.

4 Methods

4.1 Data and search

Two searches of the Web of Science Core Collection (Clarivate) were run, the first on 31 December 2025 and the second on 6 October 2026, and the second was merged with OpenAlex. The first search returned 11,174 records and served to establish the profile structure of the field, that is, the recurring types of study this literature contains, defined by sampling region, sample type, indicator type, scientific question and the Sustainable Development Goals addressed. These profiles were derived from the titles and abstracts of the first-search corpus by clustering, and DeepSeek Pro was then used to extract the key information they carry and to consolidate the category options that this paper uses throughout. The second search returned a further 42,063 records, and the same concept blocks run against OpenAlex returned 27,014. After de-duplication on DOI and normalised title 49,189 remained, of which 1,181 fell outside 1980-2025. A decision model then judged each record's relevance to palaeoenvironmental archive research, excluding 16,044 that lay outside the field and leaving the analysed corpus of 33,145. Retrieval and screening follow a PRISMA-style flow (Table S7; Figure S8). Economic data are taken directly from the World Bank World Development Indicators, series NY.GDP.PCAP.PP.KD (GDP per capita in constant 2021 PPP international dollars; retrieved 1 October 2026). The 95 countries with a World Bank series are retained and the 100 without one are excluded; no economic value was imputed, and the AI models were applied only to bibliographic metadata.

4.2 Field extraction

Every record of the merged corpus was then labelled with Jev (TypeSafe System One), a decision model that returns a calibrated distribution over a fixed option set. The option set is the one consolidated from the first-search profile clustering, and Jev assigned each record to those fixed options across ten fields, covering article type, archive carrier at major and sub-class level, proxy

indicator, scientific question, Sustainable Development Goal, sampling location and an in-scope judgement. Because the model must choose within the option set, every field is populated for every retained record; completion therefore carries no information about accuracy. Accuracy was assessed separately on a stratified manual sample and a held-out test set. The full extraction prompt and validation are in the Supplementary Information. Two AI tools were used in this workflow. DeepSeek Pro extracted the key information carried by the first-search article profiles and consolidated the category options, and Jev (TypeSafe System One) then labelled every record of the merged corpus. The authors reviewed and validated both sets of labels against a stratified manual sample and a held-out test set.

4.3 Classification and analysis

Countries are cross-classified on fractional publications and sampling-location articles. Publication credit uses fractional counting, the convention of the U.S. National Center for Science and Engineering Statistics (NCSES), dividing each paper among the countries of its authors. Sampling credit is assigned to the countries named as sampling locations in the abstract. Each dimension is split at its own P25 and P75 over countries with a non-zero value, giving 26 high-high and 39 double-zero countries. Concentration is measured by the Gini coefficient with 2,000 bootstrap replicates. Scaling is estimated by ordinary least squares on country-year observations, by a mixed model with country-level random slopes, and by a Tweedie generalised additive model for the relation between publication and sampling. Change points were located by binary segmentation with the number of segments selected by BIC, and a five-layer network traces each record from country class through archive, indicator and question to goal. The detailed specification of every test, including variable definitions, model selection and robustness checks, is given in the Supplementary Information.

Code and data availability

The bibliographic corpus itself cannot be redistributed because of the Web of Science licence. The derived record-level annotation table, the complete search string, the extraction prompt and the manual validation sample are archived on Zenodo at https://doi.org/10.5281/zenodo.23294431 under a CC BY-NC-SA 4.0 licence. The analysis code is available at https://github.com/honghlaries/natural-archive-bibliometrics under an MIT licence.

Author contributions

Lingli Fan: Conceptualization, Investigation, Software, Formal analysis, Writing - original draft.

Yiming Lei: Validation, Writing - Review & Editing.

Yufei Chen: Validation, Writing - Review & Editing.

Qian Huang: Conceptualization, Resources, Validation, Writing - review & editing.

Hui Jia: Resources, Writing - review & editing.

Shan Jiang: Writing - review & editing.

Jingchun Liu: Writing - review & editing.

Haoliang Lu: Funding acquisition, Writing - review & editing.

Chongling Yan: Writing - review & editing.

Hualong Hong: Conceptualization, Supervision, Data Curation, Funding acquisition, Writing - review & editing.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI use

During the preparation of this work the authors used two AI tools, and both were applied to bibliographic metadata rather than to the scientific content. DeepSeek Pro was used on the corpus returned by the first Web of Science search to extract key information from titles and abstracts and to consolidate the category options that define the article profile types described in the Methods, namely sampling region, sample type, indicator type, scientific question and the Sustainable Development Goals addressed. Jev (TypeSafe System One) was then used to label every record of the merged corpus, 58,201 records in total, against those fixed options, returning archive carrier at major and sub-class level, proxy indicator, scientific question, SDG framing, sampling location and an in-scope judgement; the 33,145 records retained after screening carry the labels analysed here. The authors reviewed and validated the extracted labels against a stratified manual sample and a held-out test set, and take full responsibility for the content of the publication. Generative AI tools were not used to create or alter any figure, image or other artwork, and were not used to draft or edit the text of the manuscript.

Financial support

This work was funded by the National Key R&D Program of China (2022YFF0803100) and the National Natural Science Foundation of China (U25A20801).

Acknowledgements

The authors thank the Web of Science Core Collection (Clarivate) and OpenAlex for access to the bibliographic records analysed in this study, and the developers of the R and Python packages used in the analysis.

References

1. May, R. M. The scientific wealth of nations. *Science* **275**, 793–796 (1997).

2. King, D. A. The scientific impact of nations. *Nature* **430**, 311–316 (2004).

3. Merton, R. K. The Matthew effect in science. *Science* **159**, 56–63 (1968).

4. Bol, T., de Vaan, M. & van de Rijt, A. The Matthew effect in science funding. *Proc. Natl. Acad. Sci. USA* **115**, 4887–4890 (2018).

5. Bhambra, G. K. Introduction: Knowledge production in global context: Power and coloniality. *Curr. Sociol.* **62**, 451–456 (2014).

6. Demeter, M. Academic knowledge production and the Global South: Questioning inequality and under-representation (Palgrave Macmillan, 2020).

7. Trisos, C. H., Auerbach, J. & Katti, M. Decoloniality and anti-oppressive practices for a more ethical ecology. *Nat. Ecol. Evol.* **5**, 1205–1212 (2021).

8. Sultana, F. The unbearable heaviness of climate coloniality. *Polit. Geogr.* **99**, 102638 (2022).

9. Bradley, R. S. Paleoclimatology (2015).

10. Kaufman et al. A global database of Holocene paleotemperature records. *Sci. Data* **7**, 115 (2020).

11. Masson-Delmotte, V. et al. Information from paleoclimate archives. In Climate Change 2013: The Physical Science Basis (2013).

12. Braconnot et al. Evaluation of climate models using palaeoclimatic data. *Nat. Clim. Change* **2**, 417–424 (2012).

13. Abram, N. J., McGregor, H. V., Tierney, J. E., Evans, M. N., McKay, N. P. & Kaufman, D. S. Early onset of industrial-era warming across the oceans and continents. *Nature* **536**, 411–418 (2016).

14. Tierney, J. E., Poulsen, C. J., Montañez, I. P., Bhattacharya, T., Feng, R. & Ford, H. L. Past climates inform our future. *Science* **370**, eaay3701 (2020).

15. Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R. & Huber, J. Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016. *Nature* **568**, 382–386 (2019).

16. Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).

17. Edwards, R. L., Chen, J. H. & Wasserburg, G. J. ^{238}U-^{234}U-^{230}Th-^{232}Th systematics and the precise measurement of time over the past 500,000 years. *Earth Planet. Sci. Lett.* **81**, 175–192 (1987).

18. Blaauw, M. & Christen, J. A. Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Anal.* **6**, 457–474 (2011).

19. Grissino-Mayer, H. D. & Fritts, H. C. The international tree-ring data bank: an enhanced global database serving the global scientific community. *Holocene* **7**, 235–238 (1997).

20. Esper, J., Cook, E. R. & Schweingruber, F. H. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* **295**, 2250–2253 (2002).

21. Fairchild et al. Modification and preservation of environmental signals in speleothems. *Earth-Sci. Rev.* **75**, 105–153 (2006).

22. Liu, T. & Ding, Z. Chinese loess and the paleomonsoon. *Annu. Rev. Earth Planet. Sci.* **26**, 111–145 (1998).

23. Aria, M. & Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* **11**, 959–975 (2017).

24. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W. M. How to conduct a bibliometric analysis: An overview and guidelines. *J. Bus. Res.* **133**, 285–296 (2021).

25. Zupic, I. & Čater, T. Bibliometric methods in management and organization. *Organ. Res. Methods* **18**, 429–472 (2015).

26. Chen, C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, *J. Assoc. Inf. Sci. Technol.* **57**, 359–377 (2006).

27. Raja et al. Colonial history and global economics distort our understanding of deep-time biodiversity. *Nat. Ecol. Evol.* **6**, 145–154 (2022).

28. Stefanoudis, P. V., Licuanan, W. Y., Morrison, T. H., Talma, S., Veitayaki, J. & Woodall, L. C. Turning the tide of parachute science. *Curr. Biol.* **31**, R184–R185 (2021).

29. Asase, A., Mzumara-Gawa, T. I., Owino, J. O., Peterson, A. T. & Saupe, E. Replacing "parachute science" with "global science" in ecology and conservation biology. *Conserv. Sci. Pract.* **4**, e517 (2021).

30. Haelewaters, D., Hofmann, T. A. & Romero-Olivares, A. L. Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South. *PLoS Comput. Biol.* **17**, e1009277 (2021).

31. Croudace, I. W. & Rothwell, R. G. Micro-XRF studies of sediment cores (2015).

32. Castañeda, I. S. & Schouten, S. A review of molecular organic proxies for examining modern and ancient lacustrine environments. *Quat. Sci. Rev.* **30**, 2851–2891 (2011).

33. McIntosh, K., Messin, L., Jin, P. & Mullan, Z. Countering helicopter research with equitable partnerships. *Lancet Glob. Health* **11**, e1007–e1008 (2023).

34. Ciarli, T. & Ràfols, I. The relation between research priorities and societal demands: The case of rice. *Res. Policy* **48**, 949–967 (2019).

35. Meijaard, E., Cardillo, M., Meijaard, E. M. & Possingham, H. P. Geographic bias in citation rates of conservation research. *Conserv. Biol.* **29**, 920–925 (2015).

36. Ford, J. D., Vanderbilt, W. & Berrang-Ford, L. Authorship in IPCC AR5 and its implications for content: Climate change and Indigenous populations in WGII. *Clim. Change* **113**, 201–213 (2012).

37. Amano, T., González-Varo, J. P. & Sutherland, W. J. Languages are still a major barrier to global science. *PLoS Biol.* **14**, e2000933 (2016).

38. Amano, T. et al. Tapping into non-English-language science for the conservation of global biodiversity. *PLoS Biol.* **19**, e3001296 (2021).

39. Neukom, R. & Gergis, J. Southern Hemisphere high-resolution palaeoclimate records of the last 2000 years. *Holocene* **22**, 501–524 (2012).

40. Hughes, A. C. et al. Sampling biases shape our view of the natural world. *Ecography* **44**, 1259–1269 (2021).

41. Martin, L. J., Blossey, B. & Ellis, E. Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Front. Ecol. Environ.* **10**, 195–201 (2012).

42. Secretariat of the Convention on Biological Diversity. Convention on Biological Diversity (SCBD, 1992).

43. Secretariat of the Convention on Biological Diversity. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity (SCBD, 2011).

44. Secretariat of the Convention on Biological Diversity. Kunming-Montreal Global Biodiversity Framework (SCBD, 2022).

45. Secretariat of the Convention on Biological Diversity. Digital sequence information on genetic resources, decision adopted at the sixteenth meeting of the Conference of the Parties to the Convention on Biological Diversity (SCBD, 2024).

46. Nature. Don’t rush rules for sharing digital genetic-sequence information. *Nature* **634**, 513–514 (2024).

47. Sett, S. et al. Harmonize rules for digital sequence information benefit-sharing across UN frameworks. *Nat. Commun.* **15**, 8745 (2024).

48. United Nations. Agreement under the United Nations Convention on the Law of the Sea on the Conservation and Sustainable Use of Marine Biological Diversity of Areas beyond National Jurisdiction (United Nations, 2023).

49. UNESCO. UNESCO Recommendation on Open Science (UNESCO, 2021).

576 50. UNESCO. UNESCO Science Report, The Race Against Time for Smarter Development (UNESCO, 2021).
577 51. Springer Nature. APC waivers and discounts. https://www.springernature.com/gp/open-science/policies/journal-
578 policies/apc-waiver-countries (2026).
579 52. Yang, W. & Huang, J. A global-local strategy for global open access. Innovation 6, 100799 (2025).
580 53. World Meteorological Organization. Systematic Observations Financing Facility (SOFF) (WMO, 2022).
581 54. UNESCO-IOC. United Nations Decade of Ocean Science for Sustainable Development (2021-2030) Implementation
582 Plan, IOC Ocean Decade Series 20 (UNESCO, 2021).
583 55. Ross-Hellauer, T. Open science, done wrong, will compound inequities. *Nature* **603**, 363 (2022).
584 56. Dangles et al. Insufficient yet improving involvement of the global south in top sustainability science publications.
585 *PLoS ONE* **17**, e0273083 (2022).
586 57. Nature Geoscience. Globalize geoscience. *Nat. Geosci.* **8**, 491 (2015).
587 58. Kumar, R., Khosla, R. & McCoy, D. Decolonising global health research: Shifting power for transformative change.
588 *PLOS Glob. Public Health* **4**, e0003141 (2024).
589 59. Dean, L., Njelesani, J., Smith, H. & Bates, I. Promoting sustainable research partnerships: A mixed-method
590 evaluation of a United Kingdom–Africa capacity strengthening award scheme. *Health Res. Policy Syst.* **13**, 81 (2015).
591 60. Abouzeid et al. Barriers to sustainable health research leadership in the Global South: Time for a Grand Bargain on
592 localization of research leadership? *Health Res. Policy Syst.* **20**, 136 (2022).
593 61. Damerow, J., Queste, B. Y., Vogt, M. & Novellino, A. Opening doors to physical sample tracking and attribution in
594 Earth and environmental sciences. *Sci. Data* **12**, 1047 (2025).
595 62. Williams, J. W. et al. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data
596 resource. *Quat. Res.* **89**, 156–177 (2018).
597 63. Diepenbroek et al. PANGAEA—an information system for environmental sciences. *Comput. Geosci.* **28**, 1201–1210
598 (2002).
599 64. International Ocean Discovery Program. Sample, Data, and Obligations Policy (IODP, 2024).
600 65. Stephenson, M. H., Cheng, Q., Wang, C., Fan, J. & Oberhänsli, R. Progress towards the establishment of the IUGS
601 Deep-time Digital Earth (DDE) programme. *Episodes* **43**, 1057–1062 (2020).
602 66. Deep-time Digital Earth. DDE Platform Global Launch at the DDE Open Science Forum held at the UNESCO HQ in
603 Paris. https://ddeworld.org/news/detail/189 (2022).

67. UNESCO. Persistent divides on science threaten to leave developing nations behind. https://www.unesco.org/en/articles/persistent-divides-science-threaten-leave-developing-nations-behind-new-unesco-report-shows (2026).
68. Past Global Changes. PAGES Data Guidelines (PAGES, 2018).
69. European Geosciences Union. EGU statement on scientific neocolonialism (EGU, 2023).
70. Ramirez-Castaneda, V., West, A. & Arias, A. A set of principles and practical suggestions for equitable fieldwork in biology. *Proc. Natl. Acad. Sci. USA* **119**, e2122667119 (2022).
71. Carroll et al. The CARE principles for Indigenous data governance. *Data Sci. J.* **19**, 43 (2020).
72. de Solla Price, D. J. Little science, big science (Columbia University Press, 1963).

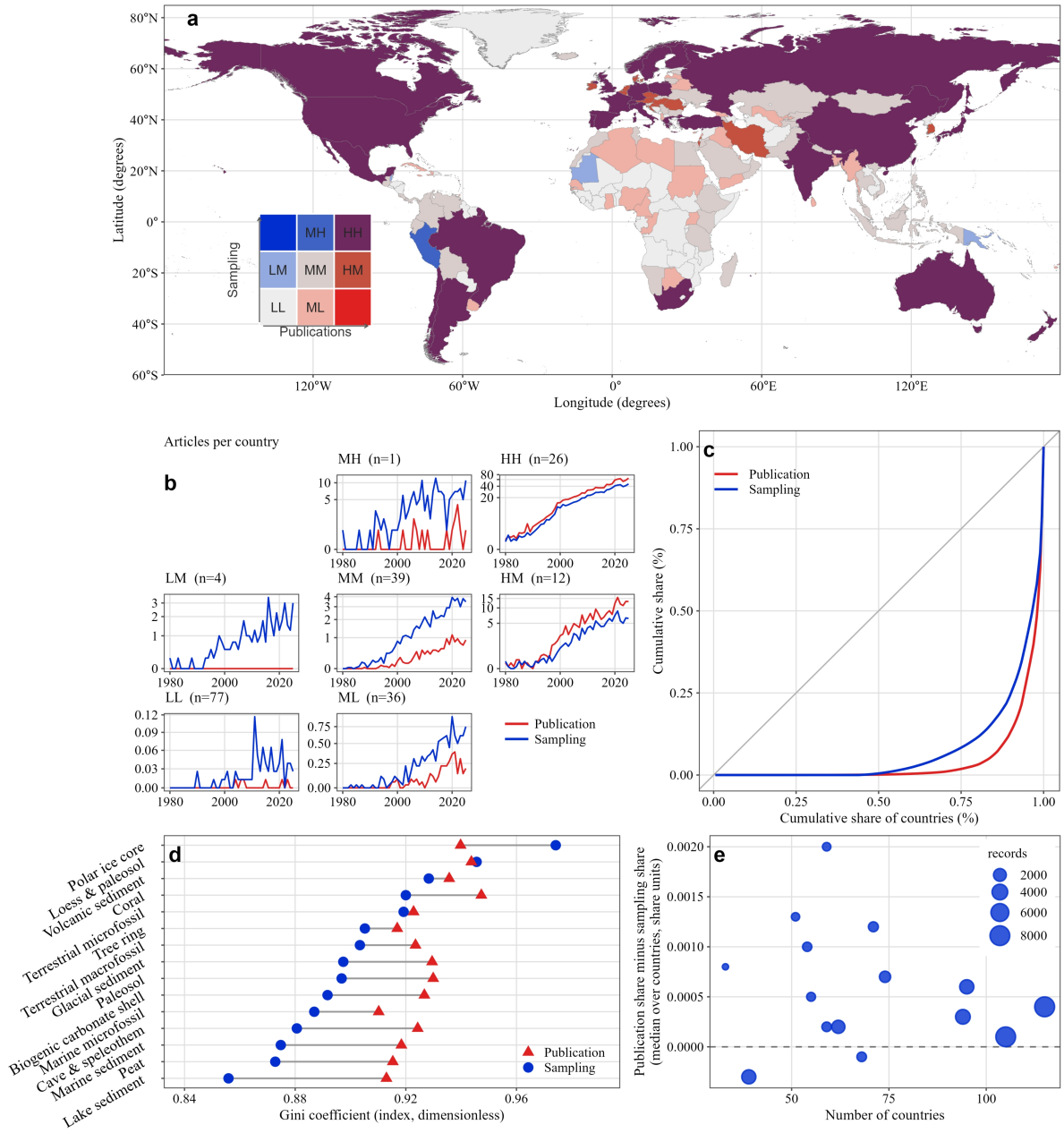


Figure 1. The global geography of production and sampling, and how the gap varies by archive sub-class.

(a) Nine-class classification of 195 countries on fractional publications and sampling-location articles; grey denotes entities outside the sovereign roster. (b) Annual publications and sampling-location articles per country within each class, plotted over the full 1980-2025 window on a common axis with years of no output filled as zero, so that the year in which each class first appears is directly

comparable across panels; the denominator is every country in the class including those with no output in that year, panels are the nine classes, titled with the class size, and empty classes (HL, LH) carry no observations. (c) Lorenz curves of fractional publications and sampling-location articles across countries, drawn on a 1:1 scale. (d) Gini coefficient of sampling-location articles (circles) and of fractional publications (triangles) across the 195-country roster, for each archive sub-class with at least 100 records; the horizontal segment gives the gap. (e) Median difference between a country's share of fractional publications and its share of sampling-location articles for the same sub-class, plotted against the number of countries in which that sub-class is sampled; point area is proportional to the number of records, and values below the dashed line mean that the publication share is smaller than the sampling share. In (e), the polar ice-core point sits low because Antarctica and Greenland are not sovereign countries and therefore fall outside the 195-country roster.

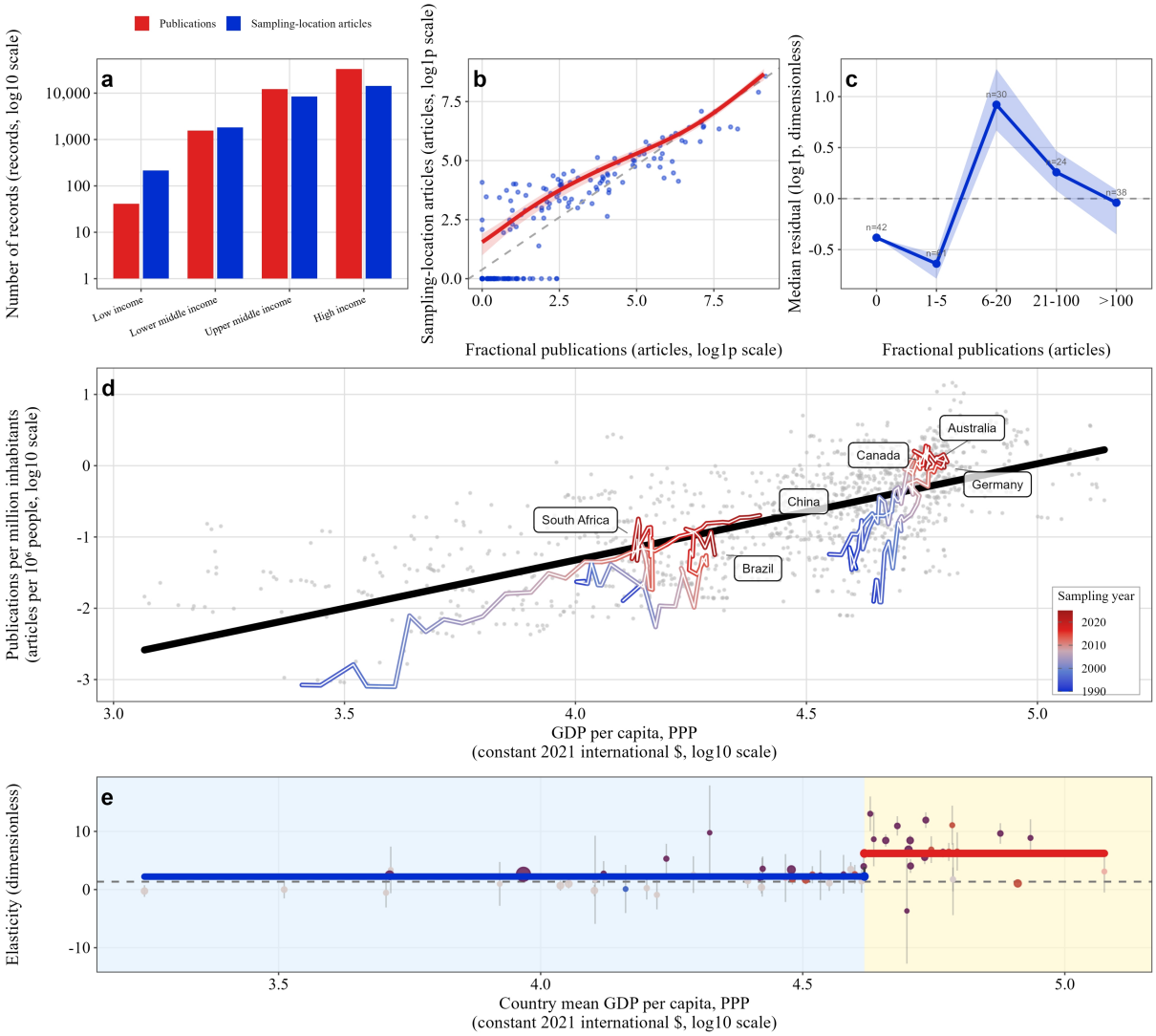


Figure 2. Income structure and the scaling step.

(a) Publications and sampling-location articles by World Bank income group (log10 scale). (b) Country-level fractional publications against sampling-location articles on log1p axes, because both counts contain zeros. The grey dashed line is a single ordinary least squares fit; the red curve is a Tweedie generalised additive model (compound Poisson-gamma, which accommodates the zero mass) with the smoothness chosen by generalised cross-validation and the shaded band giving the 95% interval of the fitted mean. The response is zero-inflated, 85 of 195 countries record no sampling, which is why the curve is fitted on the original count scale rather than on the transformed one. (c) Median residual from the ordinary least squares fit shown in (b), grouped by fractional

output; points are medians, shaded bands are 95% bootstrap intervals, the dashed line is zero and n gives the number of countries in each group. (d) Country-year observations with a World Bank PPP series (grey), with the ordinary least squares fit of publications per million inhabitants on GDP per capita in constant 2021 PPP international dollars. Coloured paths trace one representative per continent, selected as the country with the largest change in income and in publications per million among those with at least 30 annual observations, or, where no country on the continent reaches that threshold, that continent's best-covered country. China, Germany, Canada, Australia, Brazil and South Africa. Paths are coloured by sampling year, from blue (earliest) to red (latest). (e) Country-specific elasticities against mean income; the two background bands and horizontal fit lines are colour-coded by regime, blue below the breakpoint of 41,458 PPP international dollars and red above it, where the elasticity jumps by 3.96 (HC3 $p = 0.0009$). Bars are 95% confidence intervals and the dashed grey line is the pooled exponent from (d).

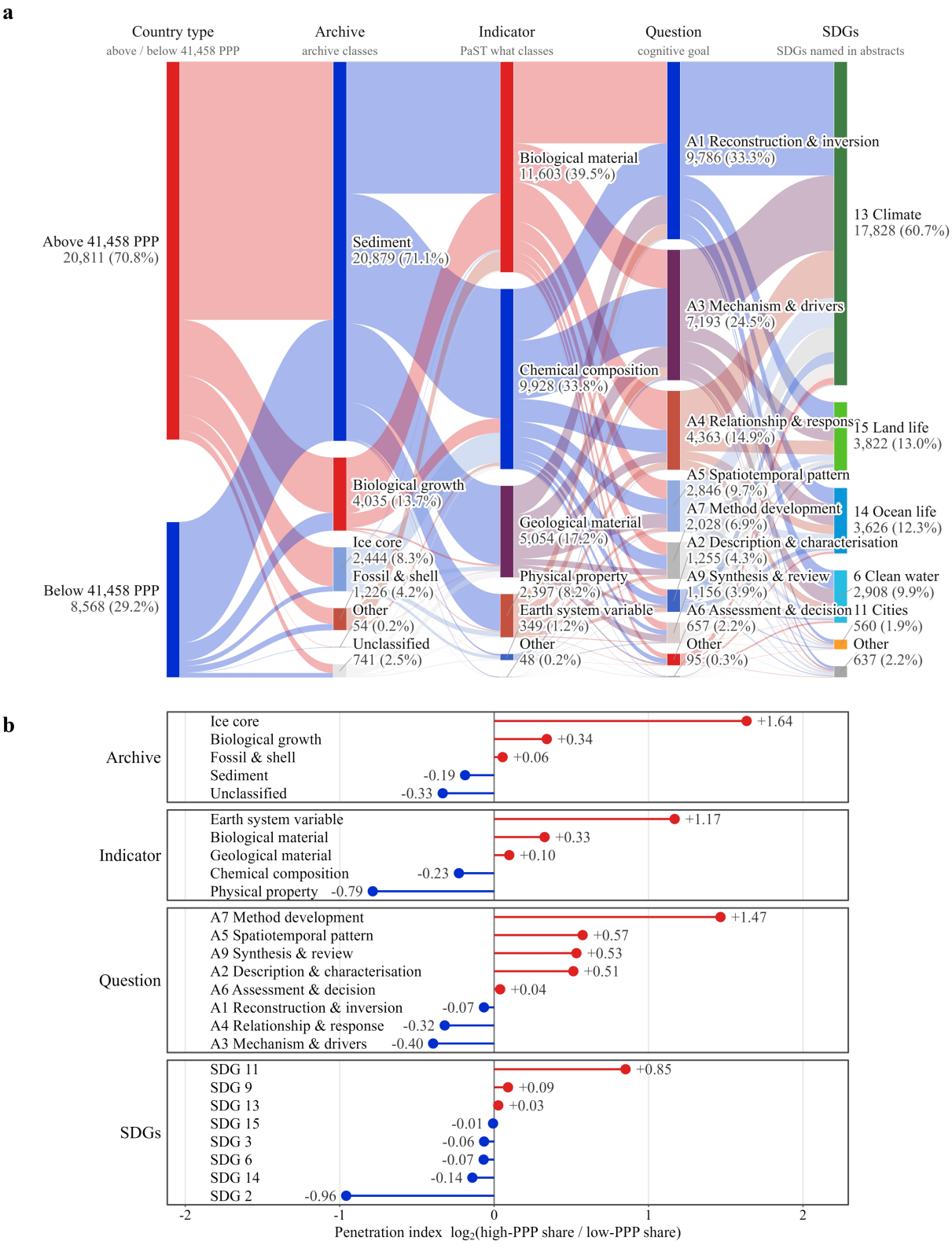


Figure 3. The five-layer knowledge network, and how far national capacity penetrates it.

(a) Flow runs from country type through archive sub-class and indicator class to cognitive goal and Sustainable Development Goal, for the 29,379 records whose first-author country has a complete GDP per capita series. Countries are split at the elasticity breakpoint of 41,458 PPP international dollars identified in Figure 2, so the source layer holds 33 countries above the breakpoint, carrying 20,811 records or 70.8% of the flow, and 62 below it; the 100 countries without a GDP series are excluded, and the retained set covers 95.4% of fractional publications. Each record contributes a total weight of 1, split evenly across the labels it carries in every multi-label layer, so the five layers conserve the record count and their widths are directly comparable. Node labels give flow in records and each node's share within its layer, and categories below 1% of a layer are merged into Other. (b) Penetration index of each node, defined as log2 of the ratio between the share it holds among records above the breakpoint and the share it holds below it. Nodes are grouped by layer and ordered by the absolute value of the index within each group, and only nodes carrying at least 0.5% of their layer are shown. Ice core at +1.64 and method development at +1.47 mark the clearest over-representation, and both concern means rather than ends. The Sustainable Development Goal layer has the smallest total variation distance of the four, 0.022 against 0.103 to 0.120 elsewhere, but that aggregate is set by SDG 13, which carries 61% of the flow and differs between the two groups by 1.14 percentage points with p = 0.071. The smaller goals do diverge, SDG 11 being over-represented 1.8-fold with 203 additional records and p = 3.6e-08, and SDG 14 under-represented by 256 records with p = 0.004, while both groups place more than 95% of their output on SDGs 13, 15, 14 and 6.

Supplementary Information

Who reads the Earth's archives? Geography, Income, and the Uneven Evidence Base for a Shared Planet

Lingli Fan [a], Yiming Lei [a], Yufei Chen [a,b], Qian Huang [d*], Hui Jia [e], Shan Jiang [f], Jingchun Liu [a], Haoliang Lu [a], Chongling Yan [a], Hualong Hong [a,b,c*]

Affiliations

a. Key Laboratory of Ministry of Education for Coastal and Wetland Ecosystems, Xiamen University, Xiamen 361102, China

b. Coastal and Ocean Management Institute (COMI), Xiamen University, Xiamen 361102, China

c. National Observation and Research Station for the Taiwan Strait Marine Ecosystem (Xiamen University), Zhangzhou 363305, China

d. Institute of Geosciences, University of Mainz, Johann-Joachim-Becher-Weg 21, Mainz 55128, Germany

e. School of the Environment and Safety Engineering, Jiangsu University, Zhenjiang 212013, China

f. State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China

Author email addresses and ORCID

Lingli Fan (a) — 33120251152187@stu.xmu.edu.cn

Yiming Lei (a) — 33120241152277@stu.xmu.edu.cn

Yufei Chen (a,b) — 33120241152351@stu.xmu.edu.cn

Qian Huang (d) * — qhuang01@uni-mainz.de

Hui Jia (e) — 1000005403@ujs.edu.cn

Shan Jiang (f) — sjiang@sklec.ecnu.edu.cn

Jingchun Liu (a) — liujingchun@xmu.edu.cn

Haoliang Lu (a) — luhl@xmu.edu.cn, 0000-0003-0720-3137

Chongling Yan (a) — ycl@xmu.edu.cn

Hualong Hong (a,b,c) * — honghl@xmu.edu.cn, 0000-0003-4284-8096

Supplementary text

S1. Search string

TS=(chronology OR chronological OR "sediment core$" OR "lake core$" OR "ice core$" OR "tree ring$" OR dendrochronolog* OR "coral skeleton*" OR "speleothem*" OR stalagmite* OR stalactite* OR "cave deposit*" OR "peat core$" OR varve* OR "laminated sediment*" OR loess OR "packrat midden" OR coprolite* OR "mollusk shell$" OR otolith* OR "ground ice" OR "permafrost core$")

AND TS=(proxy OR proxies OR surrogate OR biomarker$ OR isotope$ OR "environmental DNA" OR "ancient DNA" OR "sedimentary ancient DNA" OR sedaDNA OR eDNA OR aDNA OR "geochemi*" OR "element* ratio" OR pollen OR phytolith* OR diatom* OR foraminifera* OR "biogenic silica" OR BSi OR ostracod*)

AND DT = ("Article" OR "Early Access" OR Review)

S2. Extraction summary

Article profile types were derived from the titles and abstracts of the first Web of Science search (31 December 2025) by clustering on sampling region, sample type, indicator type, scientific question and Sustainable Development Goals addressed, and DeepSeek Pro was used to extract the key information carried by those profiles and to consolidate the category options. Jev (TypeSafe System One) then labelled every record of the merged corpus against that fixed option set, returning a calibrated distribution for each of ten fields, covering article type, archive carrier (major class and sub-class), proxy indicator, scientific question, SDG framing, sampling location and an in-scope judgement. Because Jev must choose from the option set, every field is populated for every record and completion is 100.00% for all ten fields by construction (Table S3); completion therefore carries no information about accuracy. Accuracy was assessed separately against a stratified manual sample and a held-out testset. Labels produced October 2026.

S3. Growth of the field and its developmental periods

Annual output rose from 26 articles in 1980 to 1,978 articles in 2025. It reached a local maximum of 1933 articles in 2021, dipped for two years, and has averaged 1,857.8 articles per year over 2021-2025, returning to a new high of 1978 in 2025. BIC-selected binary segmentation identifies three change points (2000, 2009, 2017) and therefore four periods, with 2,489 records in 1980-1999, 5,685 in 2000-2008, 8,471 in 2009-2016 and 15,319 in 2017-2025. Sediment archives dominate every period but their share declines monotonically, from 70.87% to 70.6%, while the share of labelled records the dictionary cannot classify grows from 1.29% to 2.75%. (Figure S7; Tables S2, S3, S4).

Composition changes within periods as well as between them. Sediment archives fall from 70.87% of records in 1980-1999 to 70.6% in 2017-2025; biological-growth archives rise from 9.44% to 14.82%; ice-core archives decline from 12.58% to 8.03%. Research articles dominate every period and their share slips slightly, from 91.76% in 1980-1999 to 89.12% in 2017-2025, while review articles fall from 4.38% to 3.48% and methodological papers rise steadily from 1.00% in 1980-1999 to 1.44% in 2000-2008 and 1.77% in 2017-2025. The declining review share indicates that synthesis has not kept pace with primary output. (Figure S7)

The plateau since 2017 is therefore not saturation of the field but saturation of its largest component. Because the aggregate curve is largely a sediment curve, its flattening describes the largest archive class rather than the field as a whole, and the growth that continues is in the classes whose share rises within the same periods.

The boundaries are legible in the same terms (Figure S4). The first coincides with the peak share of methodological papers (1.44% in 2000-2008 against 1.0% in 1980-1999), consistent with improvements in dating precision and the standardisation of proxy calibration preceding the large-scale adoption of new carriers[1–3]; the second, near 2009, with the expansion of community data networks and the maturation of tree-ring programmes[4,5]; and the third with the point at which total volume stops growing while composition keeps turning over.

S4. Archive types are separate sub-fields, not interchangeable carriers

Archive types are not interchangeable, because the material sets how much can be asked of it. Sediment, the least specialised carrier, spreads across chemical composition (34.1%) and biological material (34.9%). Ice cores concentrate in chemical composition (67.5%), fossil and shell assemblages in biological material (84.7%), and biological-growth archives in chemistry (19.0%) and physical properties such as ring width and density (1.2%). A carrier that can host many indicator classes supports many questions; one that preserves a single signal supports few.

That difference in breadth, rather than any judgement about scientific value, is why sediment carries 70.85% of network flow. The field's aggregate properties - its growth, its geography, its question profile - are therefore to a large extent properties of its most permissive archive, and reporting them as properties of paleoenvironmental science is a category error that the five-layer network makes visible, because the same statistics computed on ice cores alone would describe a different field.

The archive classes are also separate political economies, and the distance between them is measurable. Across sub-classes the concentration of sampling ranges from 0.8559 to 0.9742, the number of

countries in which a class is sampled from 33 to 115, and the sampling-publication wedge is present in some classes and absent or reversed in others. Tree rings, lake sediments and marine sediments are sampled by dispersed communities in more than eighty countries; polar ice cores are sampled in 39 and mostly outside sovereign territory. A single instrument aimed at natural archives would land on those two groups very differently, and that contrast is what the polar case exposes.

S5. What is measured and what is asked: indicator and question composition

Indicator composition tracks archive physics, though less simply than the literature sometimes implies (Figure S2); how the archive classes divide across cognitive goals is shown in Figure S1. Ice-core records are dominated by chemical composition, which carries 67.5% of their flow. Biological-growth archives divide between chemical composition (19.0%) and physical properties such as ring width and density (1.2%), so they are not primarily physical. Fossil and shell assemblages are led by biological material (84.7%), with chemistry second at 12.6%. Sediment archives, the least specialised carrier, spread across chemical (34.1%) and biological indicators (34.9%), which is what continuous non-destructive core scanning makes possible[6]. This is what the material permits, since a tree ring measures growth, an ice core measures atmospheric chemistry, and a fossil assemblage measures assemblage composition, because the signal each proxy carries is set by the carrier that preserves it[7,8].

Scientific questions, by contrast, do not track archive physics at all. Reconstruction and inversion takes 33.49% of network flow and mechanism and drivers 24.26% (Table S1); the network excludes records with an unresolved layer, and the same ordering holds at record level. Prediction and simulation, where societal value would arguably be highest, takes 0.33%. The two leading goals together account for 48.9% to 72.7% of flow depending on the archive class (Figure S2), a spread of twenty-four points that is far narrower than the variation in indicator composition across the same classes. Three sources are plausible and this corpus cannot separate them, as questions are set by training and by the reward structure of the field rather than by the material at hand; reconstruction is the step that makes any archive publishable, so it is a precondition of publication rather than a choice among alternatives; and prediction requires model-data fusion infrastructure concentrated in the countries that already dominate output, which would make the low share of prediction a capacity signal rather than a preference. What the data do establish is the asymmetry itself: what can be measured is set by the archive, and what is asked is set by the community.

The five-layer network is also reported under the nine-class country typology of Figure 2, which is a relative classification measuring each country against the others rather than against an absolute income threshold (Figure S6). High-high countries carry 29,560 of the 33,145 records, or 89.2% of the flow, and all twenty largest pathways originate from them. Every one of

the eleven goals with non-zero flow draws between 81.9% and 90.1% of its evidence from the twenty-six high-high countries, and the least-developed classes contribute between 0.00% and 0.12% in every case. Those twenty-six countries hold 57.5% of world population, so the evidence reaching these goals is produced roughly twenty-eight times more intensively per head in the high-high class than in the lower-middle, low-income and low-low classes combined. The two typologies agree that the evidence base is narrow; they differ in what they attribute the narrowness to.

S6. Methods: data, extraction and statistical specification

Data source and search strategy.

Records were assembled from two bibliographic sources. The primary source is the Web of Science Core Collection (Clarivate), queried on 31 December 2025 and again on 6 October 2026 with an expanded concept block and exported as field-tagged plain text; the document-type filter was kept to Article, Early Access and Review, so proceedings, corrections and editorial material are excluded. To test whether the resulting picture depends on that single index, the same concept blocks were then run against OpenAlex, which indexes a substantially larger and more multilingual body of literature. Three concept blocks were combined with AND in both sources, natural archives and chronometric carriers; proxy indicators; and global change. After de-duplication on DOI, normalised title and accession number, the union contains 33,145 records, of which 9,126 come from the Web of Science Core Collection, 11,663 from the expanded Web of Science query and 12,356 from OpenAlex. The finer type labels reported below (Method, Case Study, Data Descriptor and the like) are the article types recorded in the records themselves, not a further filter. 1,181 records carry a publication year outside the 1980-2025 window and are retained in the corpus total but excluded from every annual and period series, so the analysed time series contains 31,964 records. Research Article accounts for 30,109 records (90.84%), Review for 1293, Method for 527 and Case Study for 250. The remaining 966 records are spread over six smaller types, the largest being Comment/Reply (20) and Editorial/Letter (12), so the four named types cover 97.08% of the corpus.

Retrieval followed a PRISMA-style flow, reported in full in Table S7 and Figure S8. The original Web of Science Core Collection search returned 11,174 records; an expanded Web of Science query returned a further 42,063; and the same concept blocks run against OpenAlex returned 27,014. After de-duplication on DOI, normalised title and accession number, 49,189 unique records remained, of which 11,055 came from the original Web of Science search, 18,449 from the expanded Web of Science query and 19,685 from OpenAlex. Every record was then screened on title and abstract against a single question, does the study analyse a natural environmental archive as its primary evidence?, and 16,044 records were excluded because the words that had matched the search blocks were used in another sense, or because the material studied was not a natural archive. Screening retained 33,145

records, of which 1,181 carry a publication year outside the 1980-2025 window and are held in the corpus total but excluded from every annual and period series.

Field extraction with a large language model.

Ten fields were extracted from titles and abstracts with Jev (TypeSafe System One), a decision model that returns a calibrated probability distribution over a fixed option set, article type, archive carrier (major class and sub-class), proxy indicator, scientific question, SDG framing, sampling location and an in-scope judgement. Because the model must choose from the option set, every field is populated for every record; completion is therefore 100.00% to 100.00% across the ten fields by construction (Table S5) and carries no information about accuracy. Accuracy was instead assessed against a stratified manual sample and against a held-out decision testset (Supplementary Methods S2). A separate screening pass applied the same model to a single question, is this article about a natural environmental archive?, and its verdicts define the corpus reported here (.

Classification of countries, archives and indicators.

Countries are cross-classified on fractional publications and sampling-location articles. Publication credit is assigned by fractional counting, the convention used by the U.S. National Center for Science and Engineering Statistics (Table S6), in which each paper is divided among the countries of its authors in proportion to how many authors each country contributes, so that a paper with two authors in one country and one each in two others awards half the credit to the first and a quarter to each of the others. Whole counting, which gives every participating country a full point, would inflate countries that collaborate widely, while crediting only the first author would systematically undercount countries that take part as co-authors. Each dimension is split at its own P25 and P75, computed over the countries with a non-zero value on that dimension (153 countries for publications, 110 for sampling); values at or below the P25 form the low band, values between the P25 and the P75 the middle band, and values above the P75 the high band. A country with no publication therefore falls in the low band without entering the quantile calculation, which prevents the breaks from being dragged towards zero by the 53 countries that have neither a publication nor a sampling-location article in the corpus. Sampling-location articles are counted at the record level after normalising the free-text sampling description to standard country names; this metric reproduces the archived country table exactly for all 195 countries (sum of absolute differences = 0). Archive classes follow an explicit dictionary of 18 sub-classes mapped to five top-level classes; records matching no sub-class are reported as unclassified rather than forced into a class. The same dictionary carries an intermediate level of 19 archive sub-classes;. Indicators are mapped to the what branch of the PaST (the Paleoenvironmental Standard Terms

thesaurus, a controlled vocabulary for palaeoenvironmental proxies and their measured properties). Scientific questions are classified by cognitive goal.

Temporal segmentation and statistical analysis.

Annual publication volume for 1980-2025 was smoothed with a three-year centred moving average. Change points were located by binary segmentation, recursively choosing the split that minimises the within-segment sum of squares with a minimum segment length of five years, and the number of change points was selected over k = 0-6 by the Bayesian Information Criterion with a per-change-point penalty of 2 ln n. The BIC surface has no interior minimum at k = 2; k = 3 and k = 4 differ by less than five BIC units and we adopt k = 3. The unrounded change points it returns are 1998, 2009 and 2019; for the period summaries we report them rounded to the nearest year boundary as 2000, 2009 and 2017, giving four developmental periods. The full BIC table and a ten-configuration sensitivity analysis are reported in Table S6 and Table S7.

A logistic curve was fitted to the annual series, giving K = 2548, r = 0.114, t0 = 2015 and R^2 = 0.9816; observed output peaked at 1978 articles in 2025 and has averaged 1857.8 articles per year since 2021. The fitted capacity is 2,548 against the 1857.8 articles per year averaged since 2021, so the field sits at about 73% of its fitted ceiling; the recent plateau describes a slowing of growth rather than saturation. Three regressions describe the relation between income and output. Country-level per-capita output was regressed on GDP per capita in double-logarithmic space over all country-year observations with a World Bank series. Country-specific elasticities were then estimated as the slopes of the same double-logarithmic relation fitted within each country with at least ten annual observations, and their dependence on income was tested with a linear mixed model with country-level random slopes against a random-intercepts-only alternative. A step model, in which the elasticity is flat below a breakpoint and shifts by a constant above it, was fitted by grid search over the breakpoint on the 20th-80th percentile range of mean income, with HC3 heteroscedasticity-consistent standard errors and a non-parametric bootstrap over countries for the breakpoint interval. The same double-logarithmic fit between the two country-level dimensions was used to group countries by publication output and to test each group's median residual against zero, by one-sided Wilcoxon signed-rank and sign tests with bootstrap intervals. Concentration is summarised by the Gini coefficient, the Herfindahl-Hirschman index and top-N shares. The five-layer network apportions multi-label records with a 1/n weight so that every record contributes exactly one unit of flow. (Table S8) Economic data are the World Bank World Development Indicators series NY.GDP.PCAP.PP.KD (GDP per capita, PPP, constant 2021 international dollars), retrieved on 1 October 2026; the indicator page is

https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD and country series were downloaded from https://api.worldbank.org/v2/country/{ISO3}/indicator/NY.GDP.PCAP.PP.KD?format=json&date=1980:2025. The 95 countries with a series are retained and the 100 without one are excluded; no economic value was imputed, and the AI tools were applied to bibliographic metadata only.

Supplementary figures

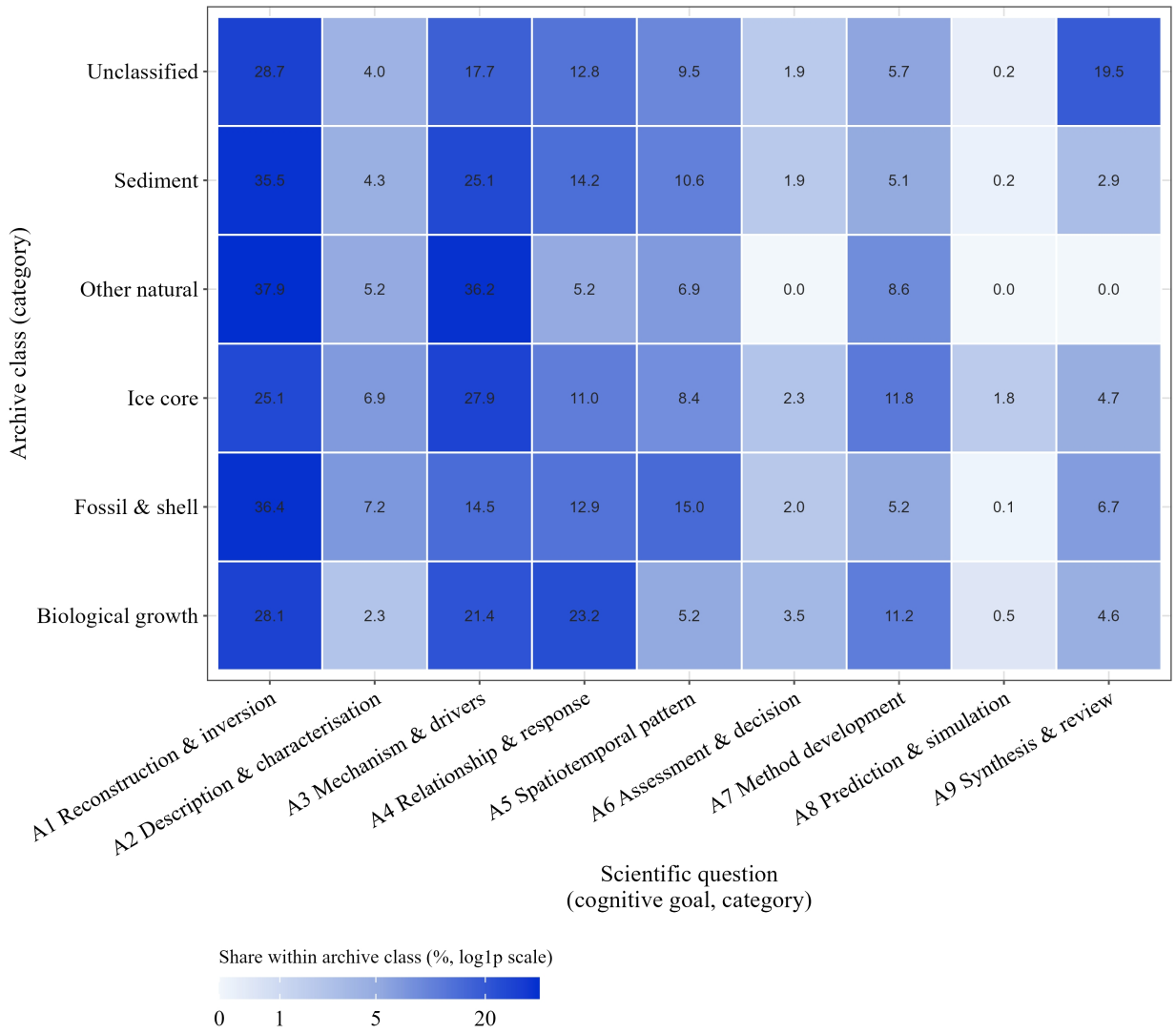
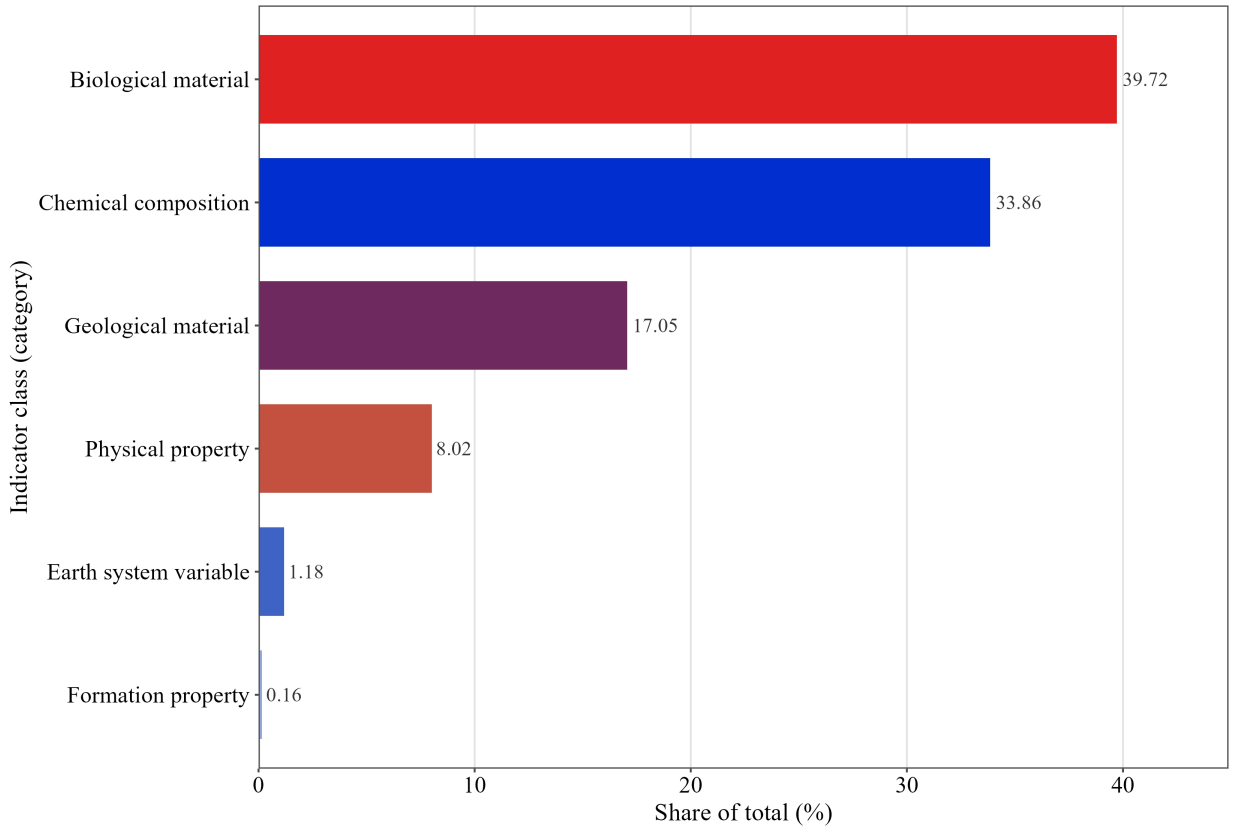


Figure S1. Share of each archive class falling into each cognitive goal.

Cells give percentages within an archive class and the colour scale is log1p.



232

233 **Figure S2. Indicator composition mapped to the what branch of the PaST thesaurus.**

234

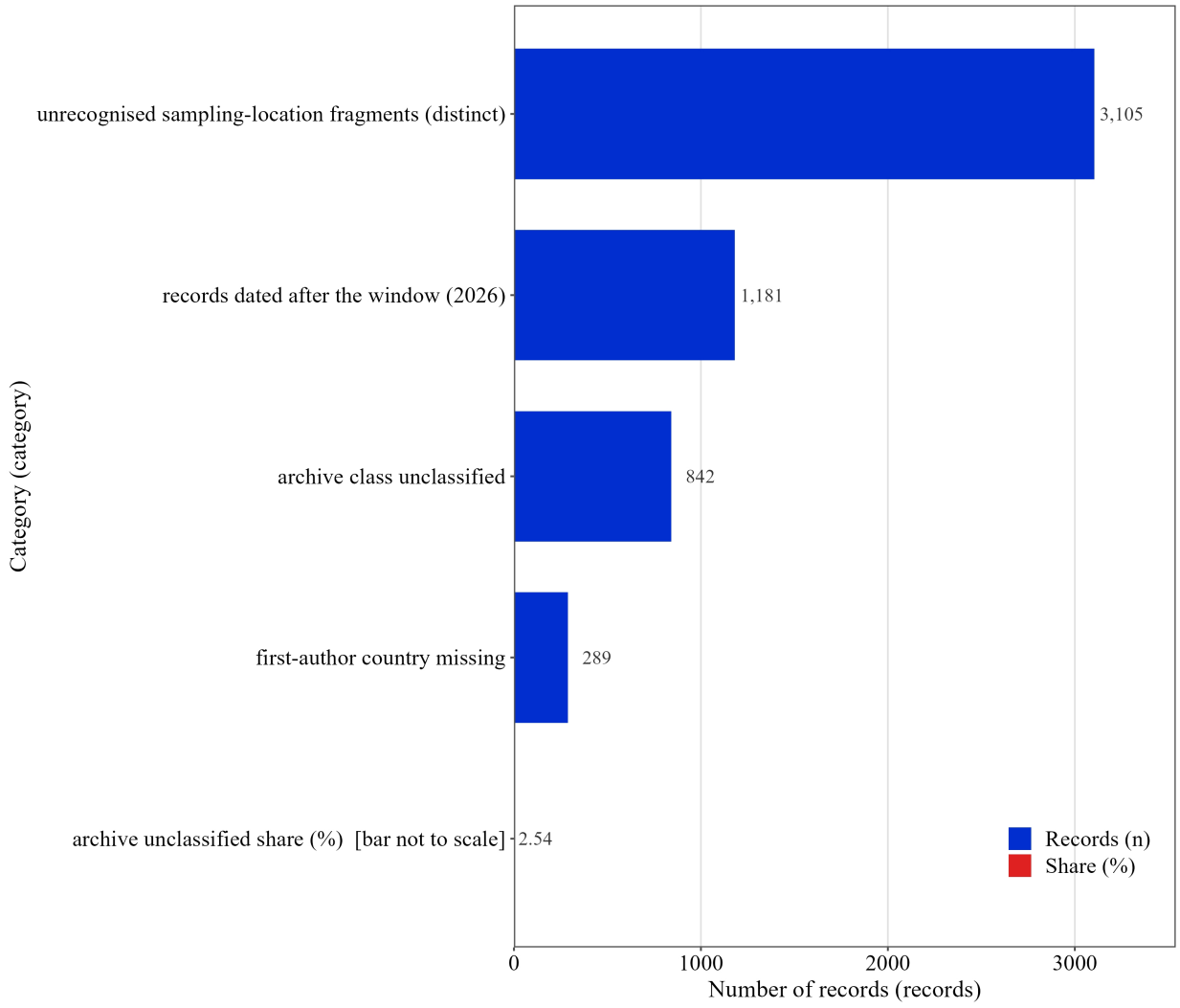


Figure S3. Data-quality profile: records with missing or unresolvable fields, and the unclassified archive residual.

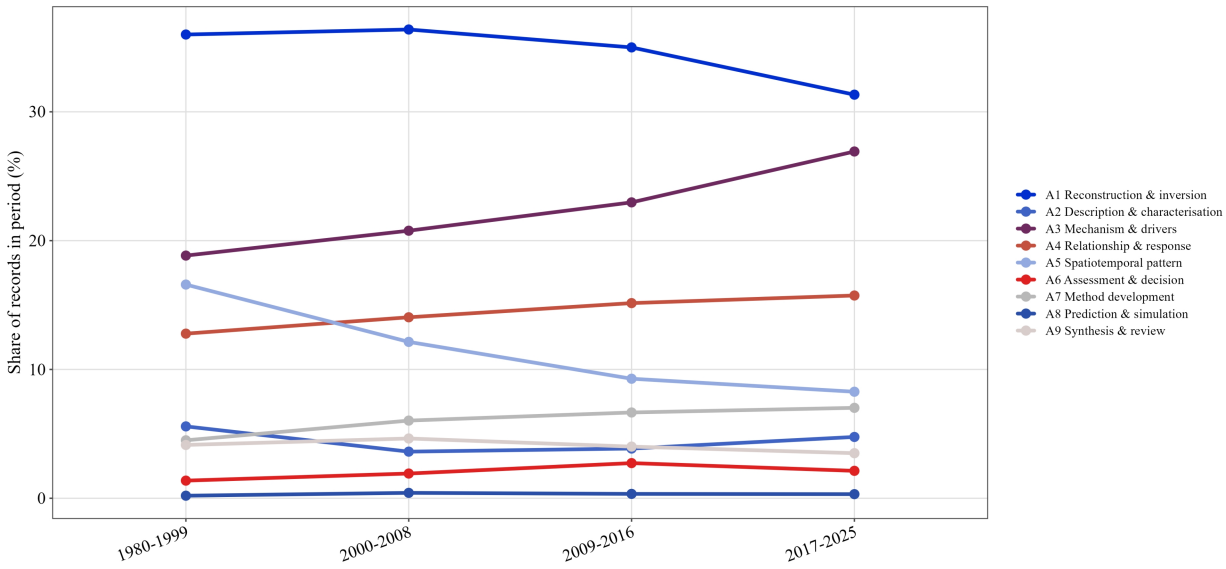


Figure S4. Cognitive goal composition across the four developmental periods.

Lines give each goal's share of the records in that period (Table S3).

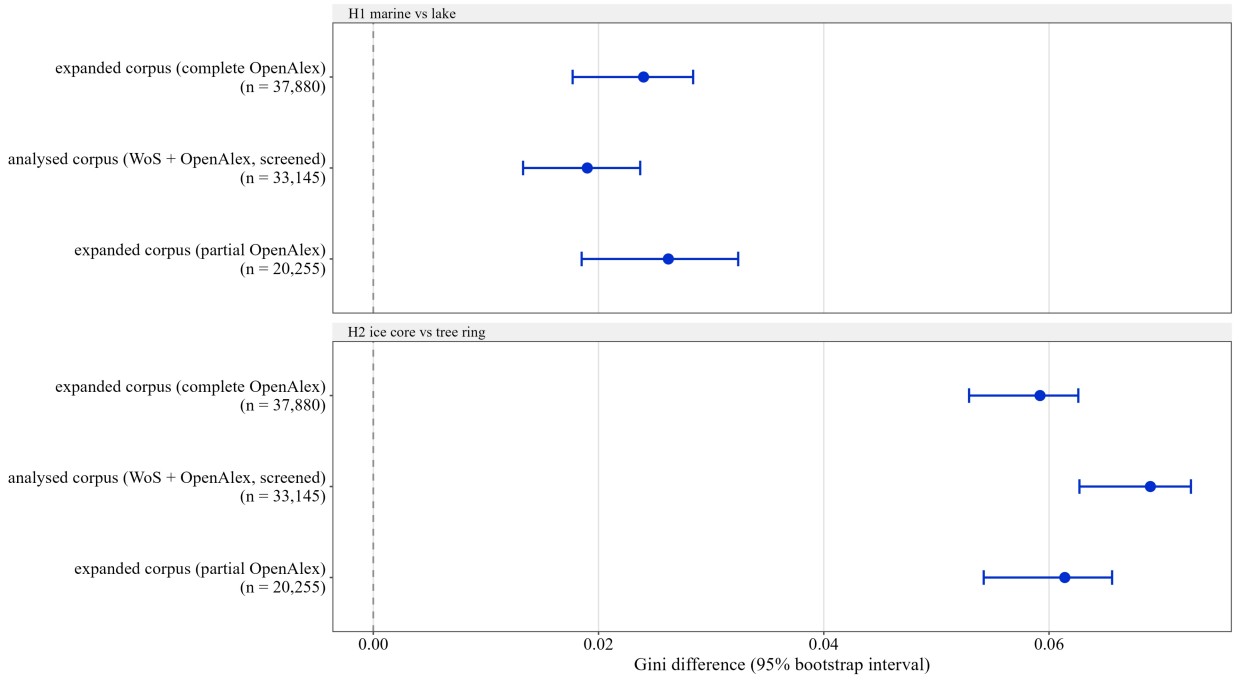


Figure S5. The two main tests evaluated on three corpus scopes, from the expanded corpus with partial OpenAlex coverage to the analysed corpus used throughout this paper.

Points are Gini differences between the two archive classes named in each panel and bars are 95% bootstrap intervals; blue marks intervals that exclude zero. Record counts for each scope appear beneath the scope label (Table S7).

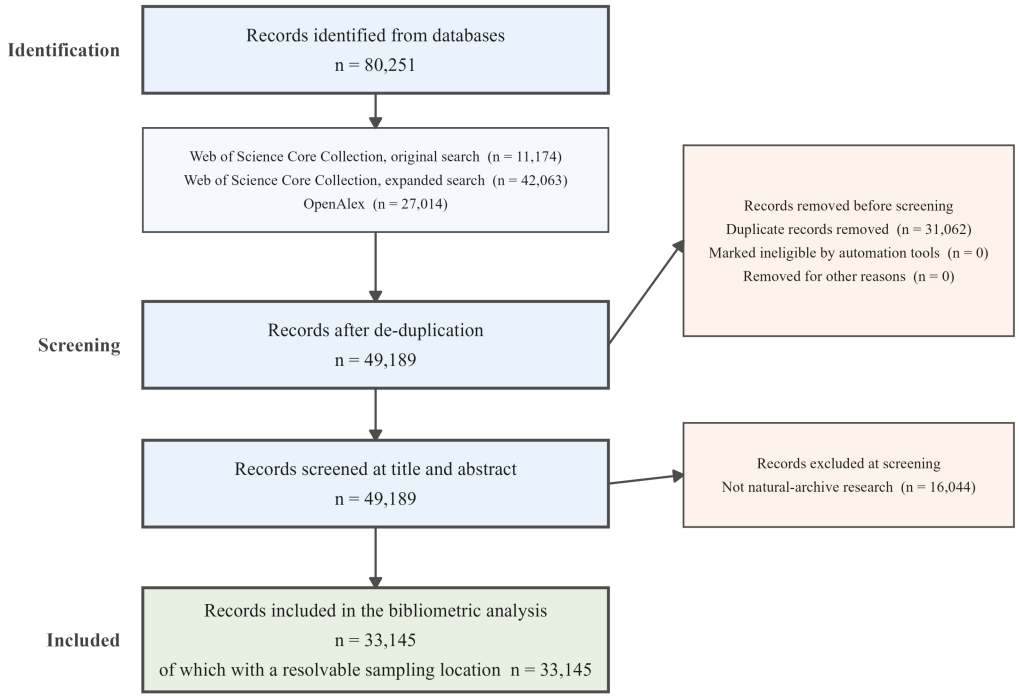


Figure S8. PRISMA 2020 flow of records from retrieval to the analysed corpus.

Boxes give the records entering and leaving each stage, arrows give the direction of flow, and the right-hand column gives the exclusions applied at each stage. Searching the Web of Science Core Collection and OpenAlex returned 80,251 records, of which 31,062 were duplicates removed before screening, leaving 49,189 for title-and-abstract screening. Screening excluded 16,044 records that were not natural-archive research, and 33,145 records entered the bibliometric analysis, all of which carry a resolvable sampling location. The stage labels follow the PRISMA 2020 statement[9].

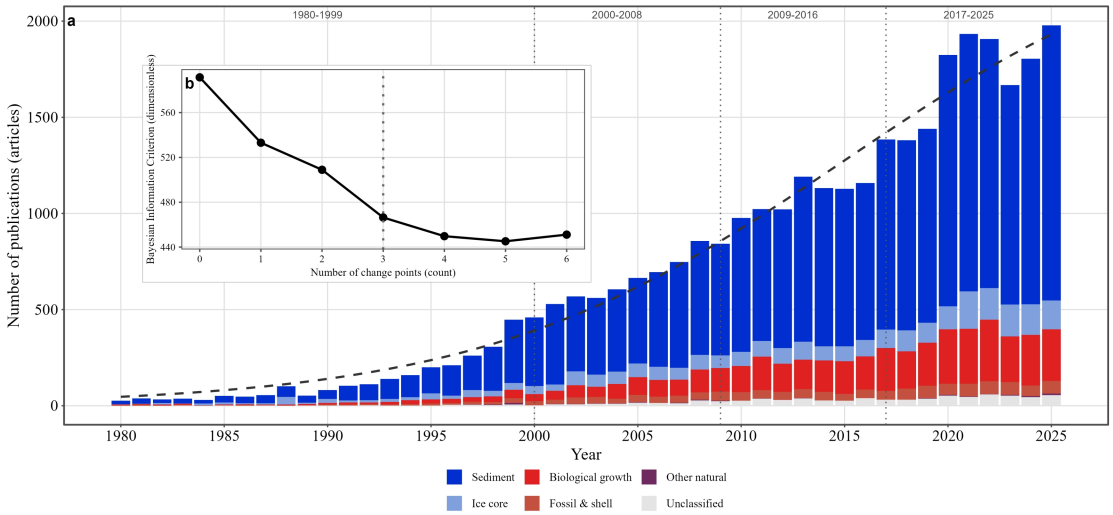


Figure S7. Growth of the field and its developmental periods.

(a) Annual output by archive class with the fitted logistic curve and the unclassified residual; (b) BIC as a function of the number of change points.

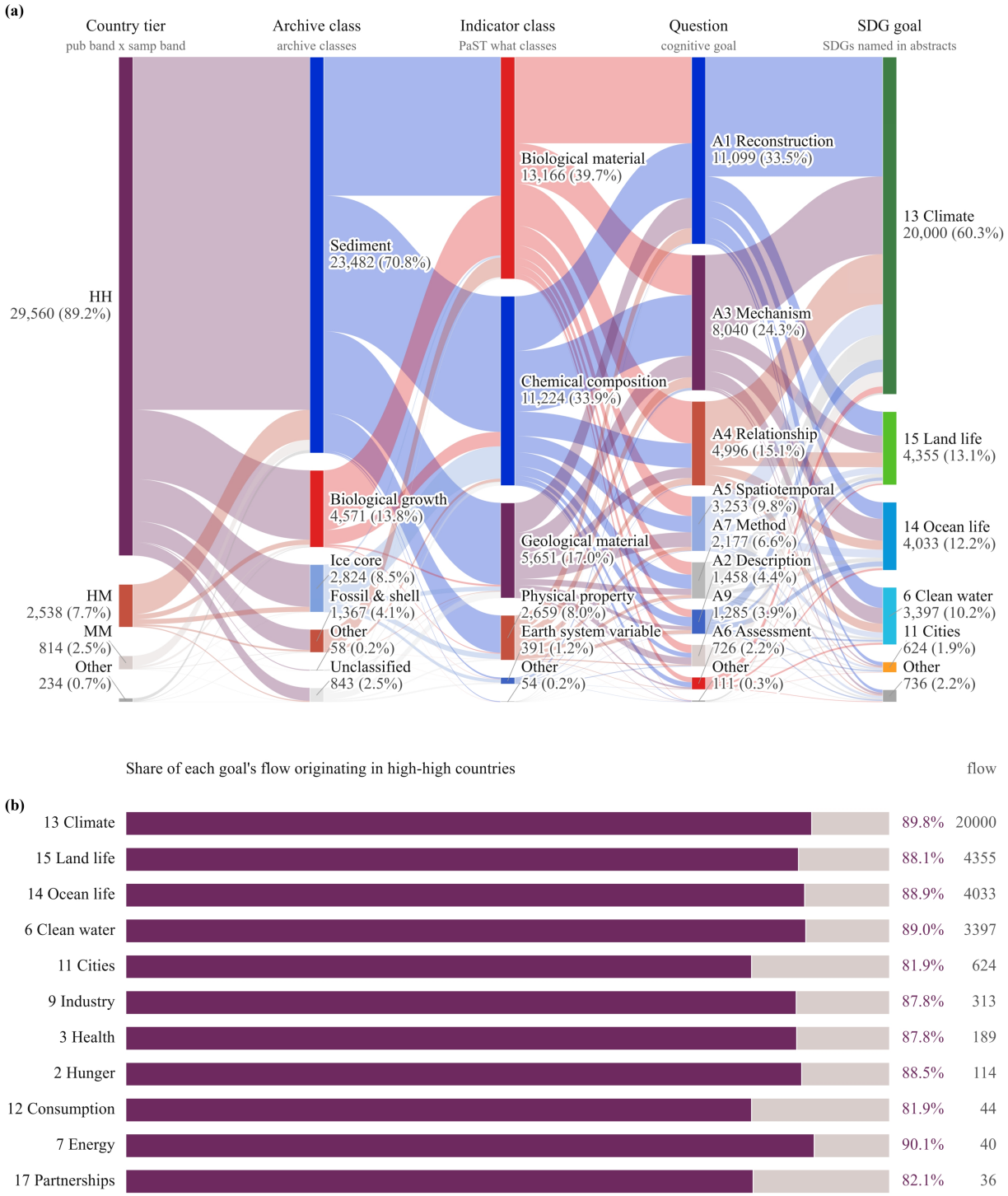


Figure S6. The five-layer network under the nine-class country typology.

(a) Flow runs from country tier through archive class and indicator class to cognitive goal and Sustainable Development Goal, for all 33,145 records carrying the five dimensions, with each record normalised to a total weight of one and split across the countries of its authors. The country tier is the nine-class typology of Figure 2, of which six classes are populated. High-high countries carry 29,560 records, 89.2% of the flow, and all twenty largest pathways originate from them. SDG 13 takes 20,000

records or 60.3% of the flow, SDG 15 13.1%, SDG 14 12.2% and SDG 6 10.2%. (b) Share of each goal's flow that originates in high-high countries, over all eleven goals with non-zero flow. Bars are drawn to a common length so that the split, not the magnitude, carries the comparison, and the flow reaching each goal is given numerically at the right. Every goal draws between 81.9% and 90.1% of its evidence from the twenty-six high-high countries, and the least-developed classes contribute between 0.00% and 0.12% in every case. This typology is relative, measuring each country's position against the others, whereas Figure 3 splits countries at an absolute income threshold; the two groupings answer different questions and are reported side by side.

Supplementary tables

Table S1. Cognitive goal composition by developmental period.

period	axis_a	records_n	share_pct
1980-1999	A1 Reconstruction & inversion	896	36.0
1980-1999	A2 Description & characterisation	139	5.58
1980-1999	A3 Mechanism & drivers	469	18.84
1980-1999	A4 Relationship & response	318	12.78
1980-1999	A5 Spatiotemporal pattern	413	16.59
1980-1999	A6 Assessment & decision	34	1.37
1980-1999	A7 Method development	112	4.5
1980-1999	A8 Prediction & simulation	5	0.2
1980-1999	A9 Synthesis & review	103	4.14
2000-2008	A1 Reconstruction & inversion	2069	36.39
2000-2008	A2 Description & characterisation	206	3.62
2000-2008	A3 Mechanism & drivers	1181	20.77
2000-2008	A4 Relationship & response	799	14.05
2000-2008	A5 Spatiotemporal pattern	690	12.14
2000-2008	A6 Assessment & decision	109	1.92
2000-2008	A7 Method development	343	6.03
2000-2008	A8 Prediction & simulation	24	0.42
2000-2008	A9 Synthesis & review	264	4.64
2009-2016	A1 Reconstruction & inversion	2965	35.0
2009-2016	A2 Description & characterisation	327	3.86
2009-2016	A3 Mechanism & drivers	1946	22.97
2009-2016	A4 Relationship & response	1283	15.15
2009-2016	A5 Spatiotemporal	786	9.28

	pattern		
2009-2016	A6 Assessment & decision	231	2.73
2009-2016	A7 Method development	564	6.66
2009-2016	A8 Prediction & simulation	29	0.34
2009-2016	A9 Synthesis & review	340	4.01
2017-2025	A1 Reconstruction & inversion	5169	31.33
2017-2025	A2 Description & characterisation	786	4.76
2017-2025	A3 Mechanism & drivers	4442	26.92
2017-2025	A4 Relationship & response	2596	15.74
2017-2025	A5 Spatiotemporal pattern	1364	8.27
2017-2025	A6 Assessment & decision	352	2.13
2017-2025	A7 Method development	1158	7.02
2017-2025	A8 Prediction & simulation	53	0.32
2017-2025	A9 Synthesis & review	578	3.5

Table S2. Scaling regressions under alternative specifications.

specification	n_obs	n_countries	slope	intercept	coefficient_A	r2	spearman
full panel (all observations)	1427	97	1.0089	-4.9352	1.161e-05	0.5276	0.7422
latest year per country (cross-section)	97	97	1.0266	-4.8878	1.295e-05	0.662	0.8142
1990-2025 panel	1403	97	1.0106	-4.9294	1.177e-05	0.5426	0.7476
below threshold (<= US$35,119)	974	93	0.8144	-4.2148	6.099e-05	0.332	0.5509
above threshold (> US$35,119)	453	27	2.3204	-11.0527	8.858e-12	0.3953	0.6051

Table S3. Stratified analysis robustness checks.

item	value
records in corpus	33145.0
records with archive_mid	33145.0
records with archive_mid and sampling country	33145.0
Unclassified records (mid level)	843.0
Unclassified: year distribution chi-square p	4.87e-89
multi-label mid records (>=2)	0.0
Gini marine, primary-label only	0.8748
Gini lake, primary-label only	0.8559
difference, primary-label only	0.019
types with n < 100 (excluded from main)	3.0
polar ice core records	2824.0
polar ice core records sampled in Antarctica	1388.0
polar ice core records sampled in Greenland	642.0
polar ice core records sampled in both Antarctica and Greenland	0.0
polar ice core records sampled outside the 195-country roster	2265.0

Table S4. The two archive-type contrasts across corpus scopes. The comparisons set marine against lake sediment and polar ice core against tree ring; difference is the Gini gap between the two classes with its 95% bootstrap interval. The analysed corpus is the screened union of Web of Science and OpenAlex used throughout this paper.

comparison	corpus	records	difference	ci_low	ci_high	ci_excludes_zero
Marine vs lake sediment	expanded corpus (partial OpenAlex)	20255	0.0262	0.0185	0.0324	True
Polar ice core vs tree ring	expanded corpus (partial OpenAlex)	20255	0.0614	0.0542	0.0656	True
Marine vs lake sediment	expanded corpus (complete OpenAlex)	37880	0.024	0.0177	0.0284	True
Polar ice core vs tree ring	expanded corpus (complete OpenAlex)	37880	0.0592	0.0529	0.0626	True
Marine vs lake sediment	analysed corpus (WoS + OpenAlex, screened)	33145	0.019	0.0133	0.0237	True
Polar ice core vs tree ring	analysed corpus (WoS + OpenAlex, screened)	33145	0.069	0.0627	0.0726	True

Table S5. Robustness of the elasticity step to output normalisation.

quantity	value
correlation log population vs log GDP per capita	-0.3657
population elasticity wrt GDP per capita	-0.6858
median country elasticity, per-million basis	2.5823
median country elasticity, raw-total basis	2.7245
breakpoint, per-million basis	41458.0
breakpoint, raw-total basis	40956.0
step at 41458, per-million basis	3.9647
step at 41458, raw-total basis	4.5349
step at 41458, population basis	0.3841
step p, per-million basis	1.43e-05
step p, raw-total basis	1.15e-06
population share of the raw-total step (%)	8.5
countries with >= 10 annual observations	56.0
panel observations	1389.0

Table S6. Mapping of archive sub-classes to major classes.

archive_mid_en	archive_class_en	note
Amber	Other natural	Amber: the workbook lists it under Other natural archives
Biogenic carbonate shell	Biological growth	
Cave & speleothem	Sediment	
Coral	Biological growth	
Ferromanganese crust	Other natural	
Glacial sediment	Sediment	
Lake sediment	Sediment	
Loess & paleosol	Sediment	
Marine microfossil	Fossil & shell	
Marine sediment	Sediment	
Paleosol	Sediment	Paleosol: same material as Loess & paleosol; the workbook assigns it to Sediment
Peat	Sediment	
Polar ice core	Ice core	
Rock glacier	Other natural	Rock glacier: no such sub-class in the workbook; assigned to Other natural archives as an ice-rock mixture
Terrestrial macrofossil	Fossil & shell	
Terrestrial microfossil	Fossil & shell	
Tree ring	Biological growth	
Volcanic sediment	Sediment	

Table S7. PRISMA flow of records from retrieval to analysed corpus.

stage	records_n
Records identified from Web of Science Core Collection, original search	11174
Records identified from Web of Science Core Collection, expanded search	42063
Records identified from OpenAlex	27014
Records after de-duplication across the three sources	49189
of which from the original Web of Science search	11055
of which from the expanded Web of Science search	18449
of which from OpenAlex	19685
Records excluded at title-and-abstract screening (not natural-archive research)	16044
Records retained for the bibliometric analysis	33145
records dated outside the 1980-2025 window	1181
records with a resolvable sampling location	33145

Table S8. Robustness of the main tests to the inclusion of reviews and syntheses.

specification	records	reviews_included	marine_gini	lake_gini	marine_minus_lake	ice_core_gini	tree_ring_gini	ice_minus_tree_ring	sampling_gini	countries_with_sampling
all records (adopted)	33145	1293	0.8748	0.8559	0.019	0.9742	0.9052	0.069	0.8556	121
reviews excluded	31852	0	0.8738	0.8547	0.0192	0.9741	0.9018	0.0722	0.8524	121

Supplementary References

1. Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).

2. Edwards, R. L., Chen, J. H. & Wasserburg, G. J. 238U–234U–230Th–232Th systematics and the precise measurement of time over the past 500,000 years. *Earth Planet. Sci. Lett.* **81**, 175–192 (1987).

3. Blaauw, M. & Christen, J. A. Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Anal.* **6**, 457–474 (2011).

4. Grissino-Mayer, H. D. & Fritts, H. C. The international tree-ring data bank: an enhanced global database serving the global scientific community. *Holocene* **7**, 235–238 (1997).

322 5. Esper, J., Cook, E. R. & Schweingruber, F. H. Low-frequency signals in long tree-ring chronologies
323 for reconstructing past temperature variability. *Science* **295**, 2250–2253 (2002).
324 6. Croudace, I. W. & Rothwell, R. G. Micro-XRF studies of sediment cores (2015).
325 7. Fairchild et al. Modification and preservation of environmental signals in speleothems. *Earth-Sci.*
326 *Rev.* **75**, 105–153 (2006).
327 8. Castañeda, I. S. & Schouten, S. A review of molecular organic proxies for examining modern and
328 ancient lacustrine environments. *Quat. Sci. Rev.* **30**, 2851–2891 (2011).
329 9. Page et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.
330 *BMJ* **372**, n71 (2021).