Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Regional stock status inference shaped by source dependence in inland fisheries

John Ed-Idoko
Shanghai Ocean University
Richard Kindong
Shanghai Ocean University
Tian Siquan
sqtian@shou.edu.cn

Shanghai Ocean University
Samuel Ayeh Osei
Shanghai Ocean University

Article

Keywords: data-limited fisheries, evidence synthesis, pseudoreplication, source balancing, stock status, uncertainty, West Africa

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11183226/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Regional stock status inference shaped by source dependence in inland fisheries

John O. Ed-Idoko[1], Richard Kindong[1,2,3,4*], Tian Siquan[1,2,3,4*] & Samuel Ayeh-Osei[1]

[1] College of marine living resource sciences and management, Shanghai Ocean University, Shanghai 201306, China

[2] National Engineering Research Center for Oceanic Fisheries, Shanghai Ocean University, Shanghai 201306, China

[3] Ministry of Education Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Shanghai Ocean University, Shanghai 201306, China

[4] Collaborative Innovation Center for Distant Water Fisheries, Shanghai 201306, China

Submitting author:

edidoko.john@gmail.com

Author details

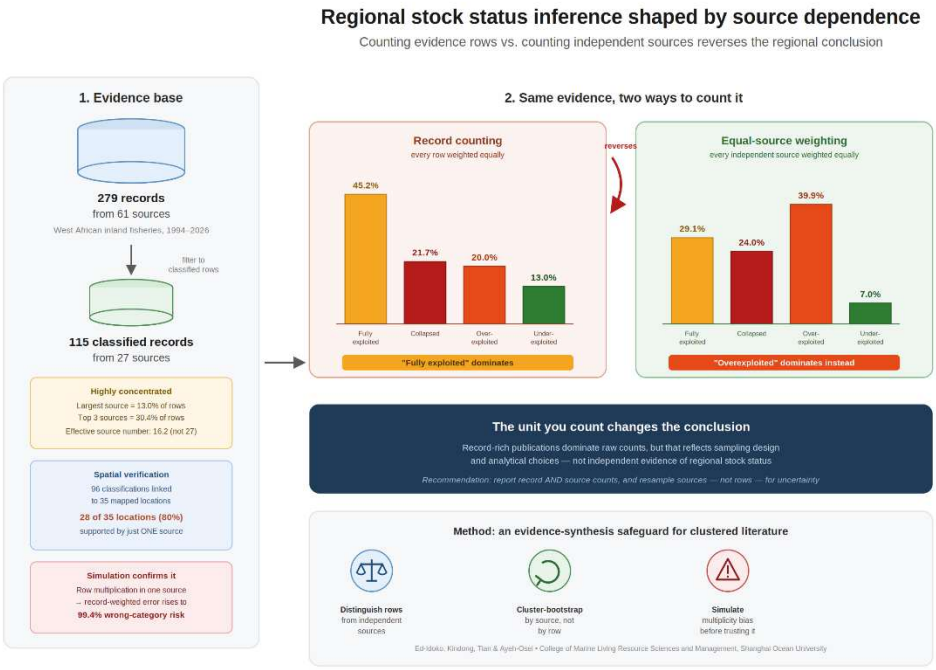
Author Number	Name	email	Orcid Number
1	John O. Ed-Idoko	edidoko.john@gmail.com	0000-0002-1311-6526
2	Richard Kindong	kindong_richard@yahoo.com	0000-0002-7793-3402
3	Tian Siquan	sqtian@shou.edu.cn	0000-0001-8693-3212
4	Samuel Ayeh Osei	samuelayehosei@gmail.com	0000-0003-0668-2607

Abstract

Freshwater fisheries evidence databases often contain many extracted records from a single publication, yet regional syntheses usually treat each row as independent evidence. Using a verified dataset of 279 records from 61 sources on West African inland fisheries, we test whether the choice of counting unit changes regional stock-status inference. Only 115 records from 27 sources carried ecological classifications, and record concentration reduced their effective source number to 16.2. Counting records, fully exploited fisheries dominated (45.2%); giving each publication equal total influence, overexploited classifications dominated (39.9% vs 29.1%), although cluster-bootstrap intervals overlapped [paired difference: insert estimate and interval]. Combined collapsed and overexploited classifications rose from 41.7% of records to 63.9% of source-balanced profiles. In simulations, multiplying rows in one category inflated record-weighted error and wrong-dominant selection (99.4% at a tenfold ratio), whereas equal-source estimates remained stable; this follows by construction and illustrates the mechanism rather than independently validating it. Indicator reporting was sparse (66.1% of complete records reported none of four indicators), and 28 of 35 mapped locations rested on a single source. Source-balanced estimands and explicit reporting of source depth are a simple safeguard for evidence syntheses with clustered observations.

Keywords: data-limited fisheries; evidence synthesis; pseudoreplication; source balancing; stock status; uncertainty; West Africa

Graphical Abstract



The infographic shows that conclusions about the status of West African inland fisheries change dramatically depending on whether analysts count individual records or independent sources.

Introduction

Freshwater ecosystems combine exceptional biodiversity with food, income and cultural functions that are disproportionately important to rural and small-scale fishing communities. Recent global assessment indicates that approximately one quarter of assessed freshwater fauna is threatened with extinction, while interacting pressures from flow alteration, pollution, habitat loss, invasive species and exploitation continue to intensify.[1–3] Small-scale fisheries contribute to nutrition, employment and multidimensional well-being at scales that are incompletely captured by conventional production statistics.[4–6]

Inland fisheries are especially difficult to assess. Rivers, floodplains, reservoirs, lakes, lagoons and wetlands contain spatially dispersed and often multispecies fisheries, while landings, effort, size structure and environmental covariates may be observed intermittently or not at all.[7–8] West African evidence illustrates this diversity: regional syntheses coexist with reservoir production models, reconstructed catches, catch-per-unit-effort analyses, length-based assessments, local ecological knowledge and single-species population studies.[9–15] These approaches are informative, but they do not supply equivalent inferential support and frequently depend on different assumptions about catchability, equilibrium, selectivity and life history.[16–20]

The difficulty is compounded when studies are assembled into evidence databases. One publication may yield a single regional record, whereas another may contribute dozens of species, water bodies or assessment outputs. Treating each extracted row as an independent replicate allows record-rich sources to dominate regional percentages, despite their shared sampling design, data stream and analytical assumptions. This is a form of pseudoreplication and is closely related to dependence among multiple effect sizes in evidence synthesis.[21–23]

Fisheries stock-assessment literature addresses analogous problems through data weighting and effective-sample-size adjustments, but these principles are rarely applied before descriptive synthesis of literature-derived status labels.[24–25]

Weighting by evidence quality is not an automatic solution. Composite scores are sensitive to coding direction, normalization and aggregation, and a source-level quality attribute copied to multiple records does not become record-specific information.[26–28] Likewise, indicator richness may reflect a single assessment package rather than independent corroboration, because catch rates, biomass ratios, exploitation rates and maximum sustainable yield can be derived from overlapping data and model assumptions.[29–31] A defensible synthesis must therefore distinguish the number of extracted records from the number of independent sources, make the estimand explicit and carry source clustering into uncertainty analysis.

Spatial coverage can create the same illusion of replication. Multiple mapped points may arise from one publication, while representative or basin-level coordinates can imply precision that the source does not contain. Georeferencing best practice requires provenance, positional uncertainty and transparent linkage between evidence records and coordinates.[32–33] Geographic research biases are particularly consequential in African freshwater systems, where data scarcity and uneven monitoring can cause evidence absence to be mistaken for ecological stability.[34–39]

Here we develop and test a verification-led framework for clustered literature evidence using West African inland fisheries as an empirical case. We ask five questions: how strongly are records concentrated within sources and countries; whether evidence limitations and indicator availability change when publications rather than records are counted; whether source balancing changes regional stock-status inference; how much independent support underlies mapped coverage; and how record multiplicity affects estimator error under controlled simulation. The goal is not to estimate the true biological status of every inland fishery in West Africa. It is to identify which regional statements are supported by independent information and to provide a transferable safeguard for evidence syntheses in which rows are nested within sources.

Results

Evidence growth masks source and geographic concentration

The verified register contained 61 publications or data resources dated 1994–2026 and 279 extracted records. Annual source output increased over the period (Kendall's $\tau = 0.491$, $P = 0.00018$; Fig. 1a), with 23 sources published during 2019–2023. Coverage nevertheless remained concentrated. Nigeria and Ghana were represented in 21 and 18 independent Study_ID–country associations, whereas Cabo Verde and Sierra Leone had spatial inventory locations but no linked literature source in the catalogue. The Gini coefficient across the 18-country spatial frame was 0.546 (Fig. 1b).

Record multiplicity varied from 1 to 53 rows per source (median 2; Fig. 1c). If record shares are interpreted as sampling weights, the concentration corresponds to an effective source number of 15.8 rather than 61. The ecological subset was more consequential: 115 classifications were nested within 27 sources, but the largest source contributed 13.0% of classified rows and the three largest contributed 30.4%. The record-weighted effective source number was only 16.2 (Fig. 1d). These diagnostics show that nominal record count substantially overstated independent information before any stock-status percentages were calculated.

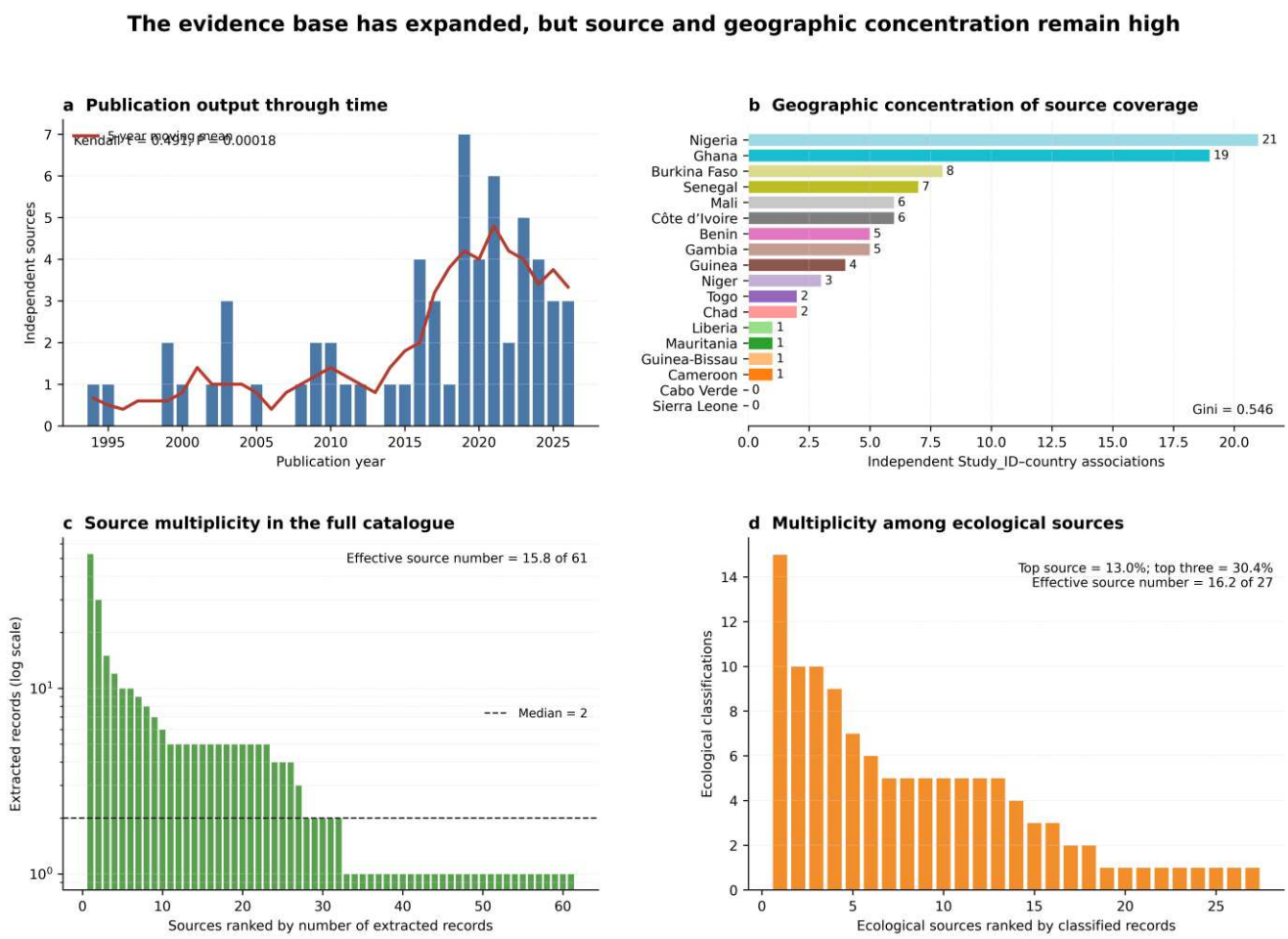


Fig. 1 | Temporal, geographic and source-level structure of the evidence base. a, Annual number of independent sources and centred five-year moving mean. b, Independent Study_ID–country associations; the Gini coefficient quantifies concentration across the 18-country spatial frame. c, Full-catalogue record multiplicity by source on a logarithmic axis. d, Ecological-classification multiplicity among the 27 contributing sources. Effective source number is the inverse Herfindahl concentration of record shares.

Evidence limitations persist after publications are counted once

Missing effort information affected 215 of 277 valid records (77.6%, 95% Wilson interval 72.4–82.1%), short time series affected 205 of 278 (73.7%, 68.3–78.6%), and multispecies aggregation affected 170 of 277 (61.4%, 55.5–66.9%; Fig. 2a). High sampling uncertainty was coded for 211 of 279 records (75.6%), high bias risk for 156 (55.9%), and low study quality for 98 (35.1%). The corresponding source-level prevalences were 76.3%, 75.0%, 76.3%, 63.9%, 57.4% and 29.5%, respectively. Among complete cases, 238 of 277 records (85.9%) and 52 of 59 sources (88.1%) had at least two of the three core design constraints.

The field labelled Confidence_Level was semantically adverse rather than favourable. It correlated strongly with Bias_Risk (Spearman's $\rho = 0.944$, $P < 0.001$), positively with Sampling_Uncertainty ($\rho = 0.415$, $P < 0.001$) and negatively with Study_quality ($\rho = -0.530$, $P < 0.001$). We therefore interpreted higher codes as greater confidence limitation. This result is a data-validity finding rather than a biological conclusion and demonstrates why field labels cannot substitute for explicit code definitions.[28,40]

Indicator prevalence and co-reporting depend on analytical grain

MSY was the most frequent of the four selected indicator fields at both analytical levels: 74 of 256 valid records (28.9%) and 22 of 55 valid sources (40.0%; Fig. 2b). B/BMSY occurred in 23.4% of valid records and 30.9% of valid sources, CPUE in 22.4% and 22.6%, and exploitation rate in 19.0% and 25.0%. Among 242 complete records, 160 (66.1%) reported none of the four indicators; 26 of 49 complete source profiles (53.1%) reported none.

Indicators were bundled rather than independent. The greatest source-level similarity occurred between B/BMSY and MSY (Jaccard = 0.789; odds ratio = 160.0, Fisher P = 3.8e-09; Fig. 2c). Exploitation rate never appeared as an isolated assessment signal in the complete source profiles and was commonly nested within MSY- or CPUE-related packages. Indicator richness therefore measured reporting breadth, not independent confirmation of stock status.[19–20,29–31]

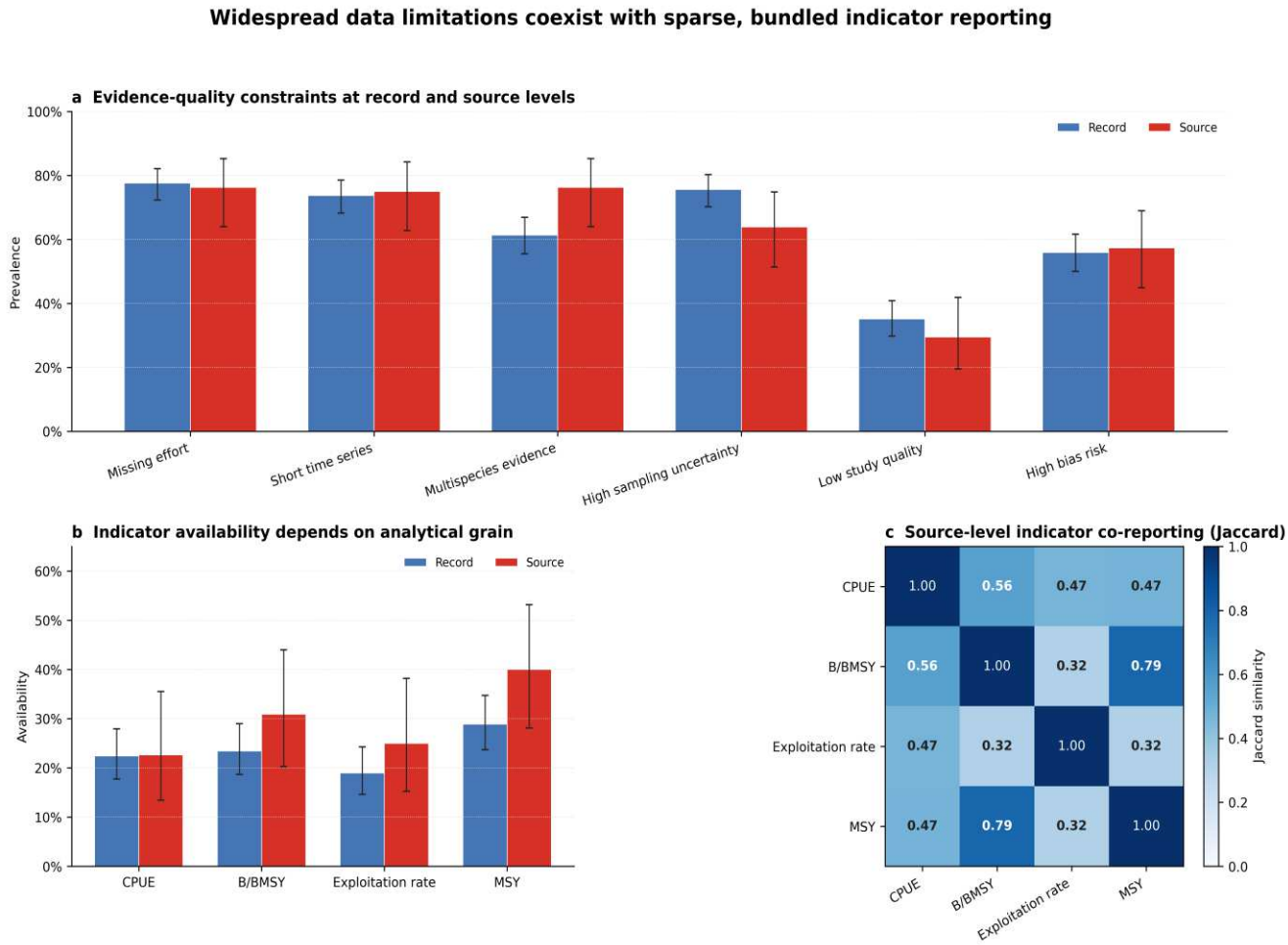


Fig. 2 | Evidence limitations and indicator structure. a, Prevalence of six quality constraints among extracted records and independent sources; error bars are 95% Wilson intervals. b, Availability of CPUE, B/BMSY, exploitation-rate and MSY fields at record and source levels. c, Jaccard similarity among source-level indicator profiles. Blank indicator values were retained as indeterminate and excluded from the corresponding denominator.

Source balancing reverses the dominant stock-status category

Of 279 evidence records, 115 from 27 sources contained an ecological classification, 11 were data-deficient and 153 contained no harmonized stock status. The ecological subset comprised 52 fully exploited, 25 collapsed, 23 overexploited and 15 underexploited records. Raw record counting therefore suggested 45.2% fully exploited, 21.7% collapsed, 20.0% overexploited and 13.0% underexploited (Fig. 3a).

Applying record-level confidence weights changed the shares to 37.9% fully exploited, 23.6% collapsed, 23.5% overexploited and 14.9% underexploited. The larger change occurred when each Study_ID received equal total influence. Overexploited classifications became dominant at 39.9%, fully exploited declined to 29.1%, collapsed accounted for 24.0% and underexploited for 7.0%. Source-balanced confidence weighting was numerically identical to unweighted source balancing because quality components were constant within every ecological source. Thus, the principal correction was for multiplicity, not repeated quality labels.

The source profiles explain the reversal. Several publications contributed many classifications concentrated in the fully exploited category, whereas adverse classifications were distributed across more sources with fewer rows (Fig. 3b). Cluster-bootstrap intervals remained wide: collapsed 9.2–40.0%, underexploited 0.0–16.5%, fully exploited 14.3–45.1% and overexploited 22.6–57.2%. Leave-one-source-out estimates did not restore the record-weighted ranking; overexploited remained within 37.6–41.5% and fully exploited within 26.4–30.2% (Fig. 3c). These intervals quantify variation among documented sources, not uncertainty about the true status of all regional fisheries.

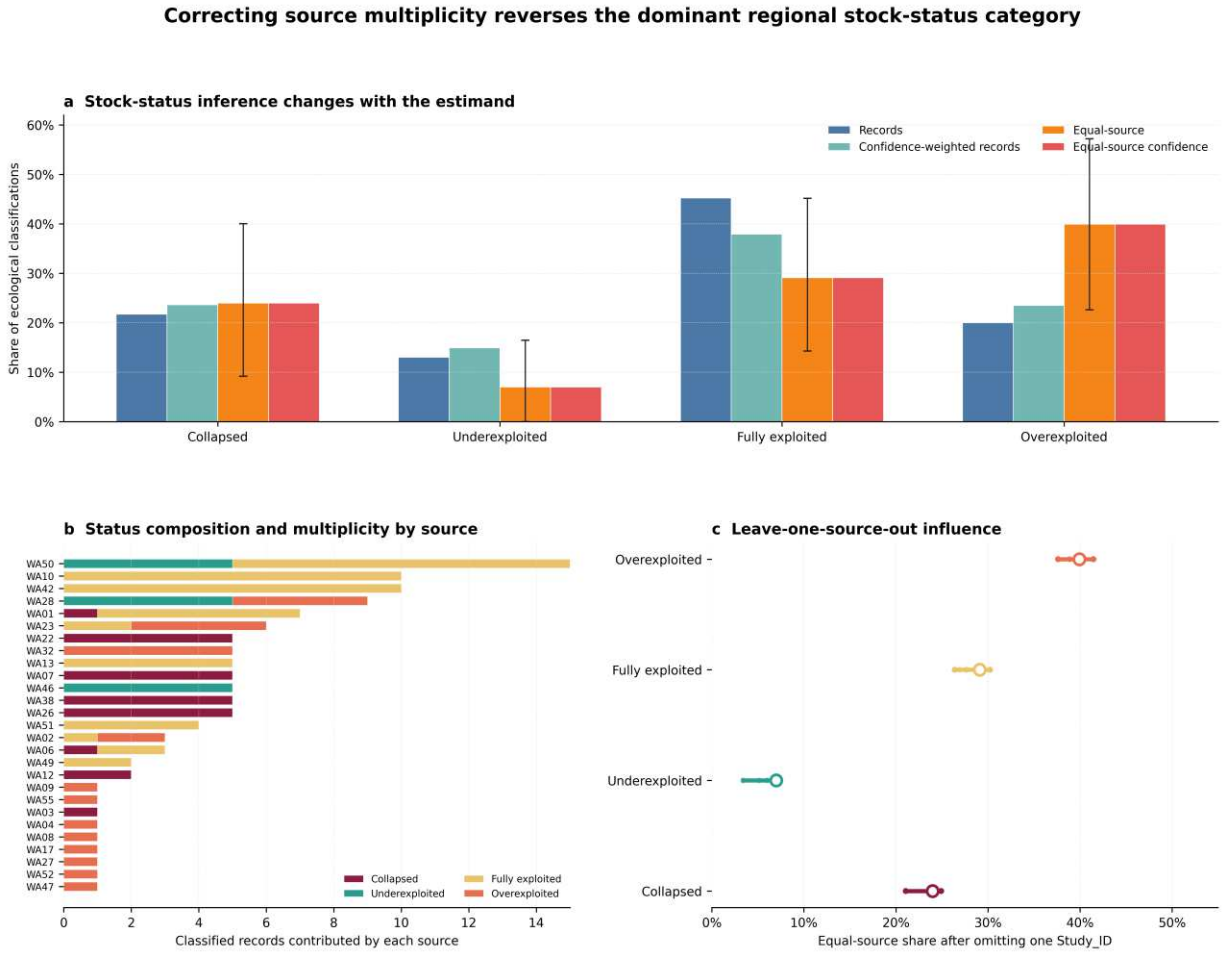


Fig. 3 | Stock-status inference under alternative analytical units. a, Regional status shares for raw records, confidence-weighted records, equal-source profiles and equal-source confidence profiles. Error bars on equal-source estimates are 95% Study_ID cluster-bootstrap intervals. b, Number and status composition of classifications contributed by each ecological source. c, Equal-source status shares after omitting each Study_ID in turn; open circles show the full equal-source estimate.

Mapped extent exceeds independent spatial replication

The repaired crosswalk linked 96 of 115 ecological classifications (83.5%) to 35 spatial locations in ten countries and 22 Study_IDs. Sixty-nine linked records used original site coordinates, two used verified sites, four used representative water-body points, ten used state-level proxies and eleven used regional or basin proxies. Coordinate provenance was retained separately from ecological confidence.[32–33]

Twenty-eight of the 35 locations were supported by one source, six by two sources and one by three sources (Fig. 4b). The map therefore contained many geographically distinct points but little repeated independent monitoring. Rivers and reservoirs dominated the linked locations, although lagoons, lakes, floodplains and basin-level contexts were also represented (Fig. 4c). Cabo Verde and Sierra Leone remained visible in the spatial inventory, but their D3 locations were not assigned ecological evidence because no corresponding D1 source existed. This distinction prevents spatial completeness from being mistaken for evidence coverage.

Country-level source-balanced profiles were retained as descriptive discovery tools rather than national assessments. Ghana and Nigeria each had at least eight ecological sources and adverse shares of 48.1% and 50.0%, respectively; leave-one-source-out ranges were 41.7–54.2% for Ghana and 42.9–57.1% for Nigeria. In contrast, several countries had apparently severe profiles based on one or two sources. Such estimates identify corroboration priorities, not precise national prevalence of overexploitation.

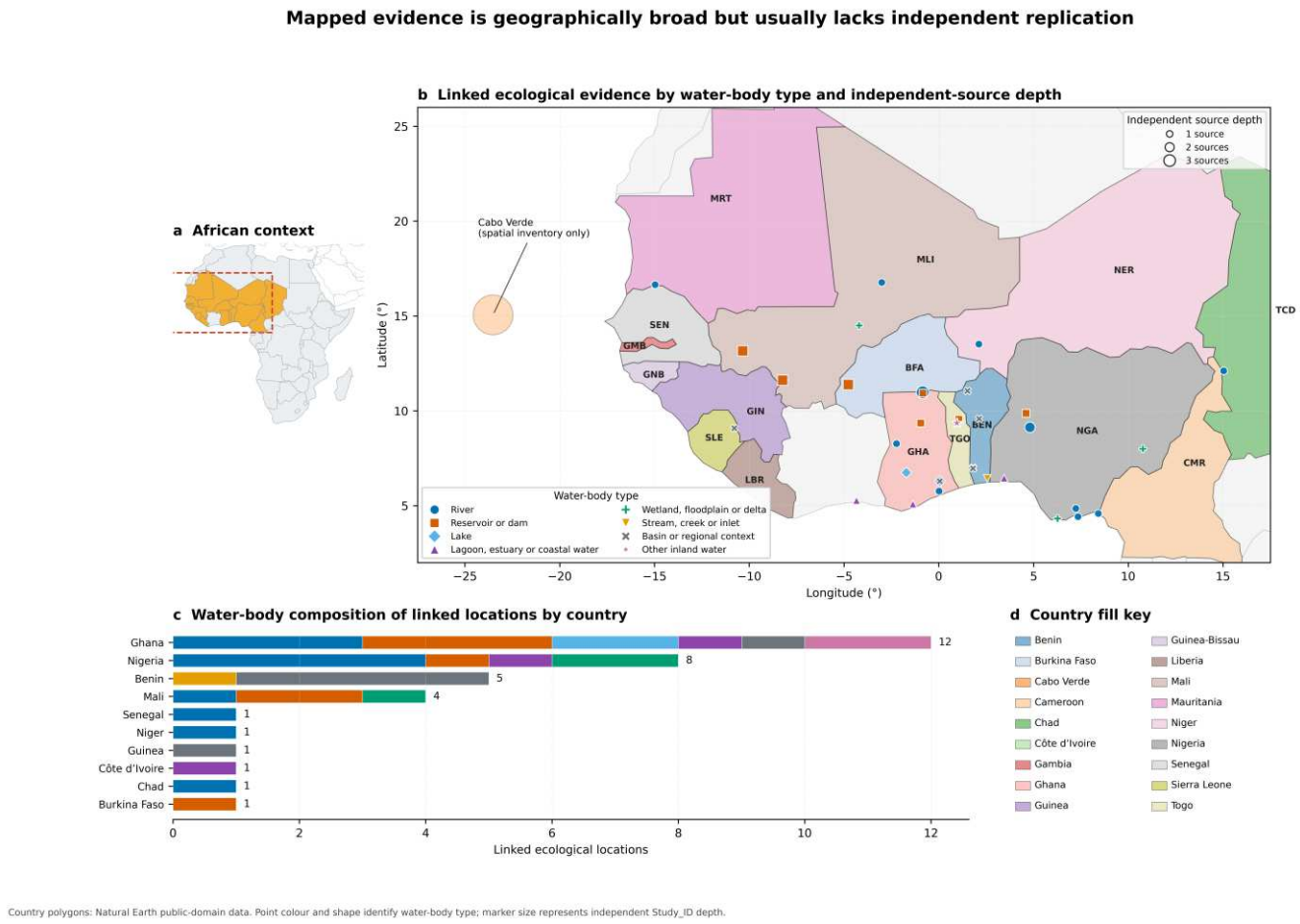


Fig. 4 | Spatial distribution and independent-source depth of linked ecological evidence. a, Africa locator. b, West Africa zoom with distinct country fills; point colour and shape indicate water-body type and marker size indicates the number of independent Study_ID values at each location. c, Composition of linked locations by country and water-body type. d, Country fill key. Cabo Verde and Sierra Leone spatial-only inventory records are not treated as ecological evidence links.

Multiplicity simulation generalizes the empirical reversal

We simulated 27-source evidence bases from the empirical equal-source status distribution and progressively multiplied the number of rows generated by fully exploited sources, while leaving the underlying source distribution unchanged. At a multiplicity ratio of 1, mean total-variation error was 0.161 for record weighting and 0.125 for equal-source estimation. When the ratio reached 2, record-weighted error increased to 0.208 and the probability of selecting the wrong dominant category increased to 74.4%. At a ratio of 10, record-weighted error reached 0.487 and wrong-dominant selection reached 99.4%. Equal-source error and dominant-category performance remained effectively stable across ratios (Fig. 5).

The simulation isolates a structural property of clustered extraction: adding rows within selected sources changes a record-weighted regional distribution even when no new independent source has been added. Equal-source estimation does not remove uncertainty about which sources are available, but it prevents record multiplicity from being misinterpreted as replication.

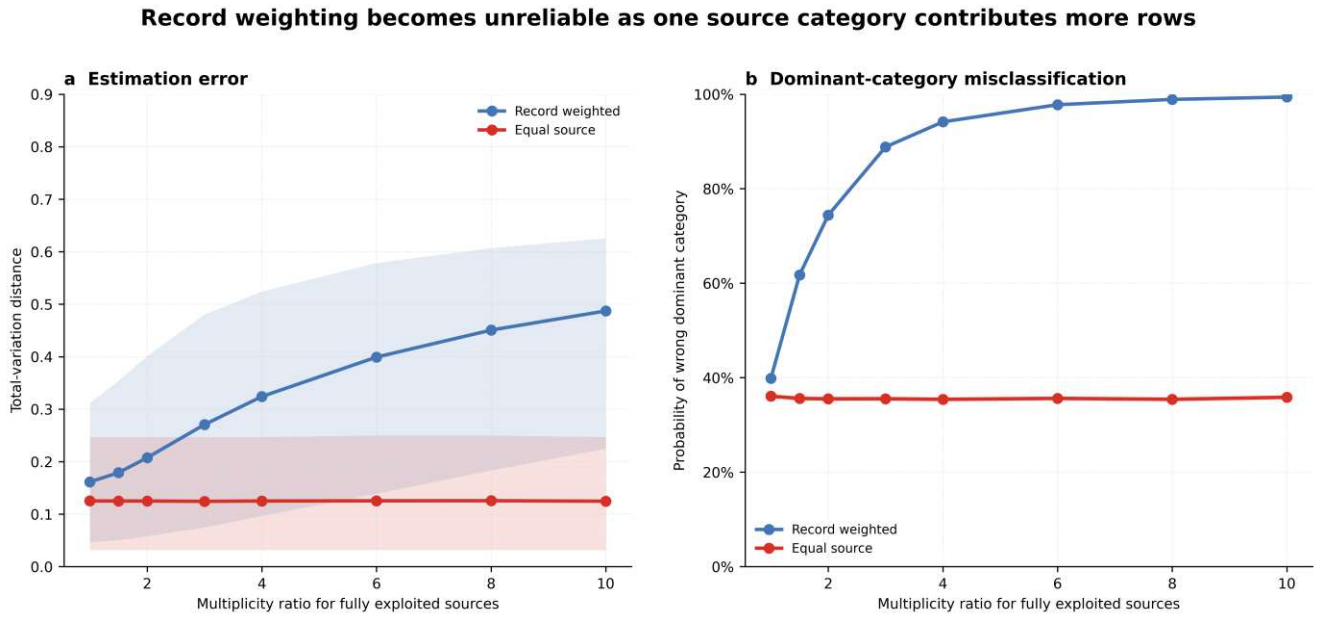


Fig. 5 | Consequences of preferential source multiplicity in simulation. a, Total-variation distance between estimated and generating source-level status distributions. b, Probability that the estimator identifies a category other than the generating dominant category. Curves show means across 30,000 simulations per ratio. Shaded regions show replicate percentiles for total-variation error and 95% Monte Carlo intervals for dominant-category probabilities.

Discussion

The central finding is that the unit used to count evidence changes the regional conclusion. Raw records identified fully exploited classifications as dominant, whereas equal weighting of independent sources identified overexploited classifications as dominant. The reversal does not show that all West African inland fisheries are overexploited, nor does it transform literature labels into a formal regional stock assessment. It shows that a conclusion supported by row counts can be an artefact of how many records individual publications contribute.

This distinction is fundamental for evidence synthesis. Record-level summaries answer how the released database is populated and are useful for assessing field completeness or computational workload. Source-level summaries answer how independent publications distribute their evidence. When multiple rows inherit the same sampling design and analytical assumptions, the latter is the defensible primary estimand for regional inference.[21–23] The effective-source calculations demonstrate the scale of the problem: 115 ecological rows represented only 27 sources, and record concentration reduced the effective source number to 16.2. Similar logic underlies data-weighting and effective-sample-size adjustments in formal stock assessment.[24–25]

Confidence weighting refined the record-weighted distribution but did not alter the equal-source result because the quality variables were constant within ecological sources. This is substantively important. Repeated source-level attributes should not be presented as if they distinguish individual species or water-body records. The semantic reversal of Confidence_Level also demonstrates that composite weighting can amplify a coding error if field direction is assumed from its label. Uncertainty partitioning and transparent, decomposable indicators are therefore prerequisites for weighting rather than optional sensitivity checks.[26–28]

The indicator analysis adds a second caution. MSY was most prevalent at source level, but more than half of complete source profiles reported none of the selected indicators and the indicators that did occur were strongly bundled. West African assessments use a wide range of methods, from reconstructed catches and reservoir yield

models to CPUE, length-frequency and local-knowledge approaches.[9–14,41–46] Counting multiple indicator fields as independent corroboration would therefore overstate evidence breadth. Indicator presence should be accompanied by method provenance, assumptions and, where possible, the original magnitude and uncertainty.

The spatial results reveal an analogous mismatch between extent and replication. The linked evidence covered diverse rivers, reservoirs, lakes, lagoons and floodplains, consistent with the heterogeneity of West African inland waters.[7,47–51] Yet four-fifths of mapped locations were represented by a single source, and a substantial minority of records required representative or broader proxy coordinates. Maps should therefore encode water-body type, coordinate provenance and independent-source depth rather than display undifferentiated points. National or basin-level management claims should not be inferred from point density alone.

The simulation establishes that the empirical reversal is not dependent on one status-coding rule. Preferential multiplication of records in a non-dominant category rapidly distorted record-weighted estimates, while equal-source estimates retained the same sampling uncertainty across multiplicity ratios. The practical implication extends beyond fisheries. Any literature database containing multiple outcomes, species, sites or scenarios per publication should report the source hierarchy and test whether headline conclusions survive source balancing.

The framework has limitations. The evidence search may not recover inaccessible, non-indexed or non-English literature, and the catalogue combines peer-reviewed articles, technical reports and data resources. Harmonized stock-status classes cannot reproduce every source-specific reference point or assumption. Source balancing gives every publication equal total influence, which is transparent but not necessarily optimal when source precision can be estimated independently. The simulation represents preferential multiplicity but not all selection processes that shape real evidence bases. Finally, point-based spatial linkage does not represent hydrographic connectivity or the spatial extent of fisheries. These limitations support conservative interpretation and motivate external validation in another data-limited regional evidence base.

For immediate practice, we recommend five minimum reporting standards: preserve an immutable source identifier for every record; report record and source counts together; define whether the target estimand concerns database rows or independent sources; propagate uncertainty by resampling sources rather than rows; and map source depth and coordinate provenance alongside locations. These steps do not replace better monitoring. They prevent existing evidence from appearing more independent and precise than it is.

Methods

Evidence base and analytical units

The study used the curated D0–D5 resource described in the companion Data Descriptor. The source register contained 61 Study_ID values and the evidence catalogue contained 279 Record_ID values. Eligible sources reported ecological status, population or stock assessment, catch and effort, fishery characteristics, environmental context or fisheries-dependent socioeconomic information for West African inland fisheries and connected transboundary systems. Marine-only and aquaculture-only evidence was excluded unless inland capture-fisheries information was separable. Search and screening documentation should be reported according to PRISMA and ROSES principles in the final submission.[52–53]

The fundamental hierarchy was Record_ID nested within Study_ID. A record represented a species-, water-body-, assessment- or country-level extraction unit. Study_ID represented the publication or data resource from which records were extracted. Analyses that described database completeness used records; analyses that

inferred literature-level patterns used sources. No publication-level duplicates were consolidated after the bibliographic audit because the corrected register treated all 61 Study_ID values as distinct source records.

Country coverage and source concentration

Annual publication counts were calculated for every year from the minimum to maximum publication year, including years with zero sources. A Kendall rank test evaluated monotonic change. Geographic coverage used unique Study_ID–country pairs after deterministic normalization of punctuation and country-name variants. The country frame comprised the 18 countries represented in D3, including zero-source countries. Concentration was summarized with the Gini coefficient. Source multiplicity was the number of D1 rows per Study_ID. Effective source number was calculated as $1/\Sigma p_i^2$, where p_i was the proportion of records contributed by source i; this equals the inverse Herfindahl concentration of record shares.

Evidence-quality profiles

Six coded limitations were examined: missing effort data, short time series, multispecies evidence, high sampling uncertainty, low study quality and high bias risk. Record-level prevalence used non-missing values. Source-level profiles used the modal value within Study_ID; ties were resolved conservatively toward the limitation state. Prevalence intervals were Wilson 95% intervals. Co-occurrence used complete cases for missing effort, short time series and multispecies evidence. The semantic direction of Confidence_Level was tested using Spearman correlations with Bias_Risk, Sampling_Uncertainty and Study_quality rather than inferred from the field name.

Indicator availability and co-reporting

CPUE, B/BMSY, exploitation rate and MSY were analysed as availability fields, not magnitudes. At record level, 1 indicated that the indicator was reported or used, 0 indicated observed absence and blank remained indeterminate. At source level, an indicator was present if any record was 1, absent if at least one record was observed and all observed records were 0, and indeterminate if all records were blank. Indicator richness was the sum of the four fields in complete profiles. Pairwise source-level co-reporting was evaluated with Fisher exact tests, odds ratios and Jaccard similarity. Because the six tests were descriptive and strongly related, exact P values are reported but no claim rests on a single significance threshold.

Stock-status estimands and confidence weights

Stock-status codes were harmonized as −1 (collapsed), 0 (data-deficient), 1 (underexploited), 2 (fully exploited), and 3 (overexploited). Ecological proportions included −1, 1, 2, and 3; data-deficient and blank records were retained as evidence states but excluded from ecological denominators.

We compared four estimands:

1. Record estimand: Each classified record received equal weight. This describes how the released database is populated but is sensitive to sources contributing many rows.

2. Confidence-weighted record estimand: The record estimand gave every classified row weight one. The confidence-weighted record estimand used $W_i = (E_i Q_i R_i)^{1/3}$, where $E_i = (r_i + 1)/5$ and r_i was the number of reported indicator fields; Q_i = Study_quality/3; and R_i was the minimum of reverse-coded Bias_Risk and reverse-coded Confidence_Level. The positive floor in E prevented records supported by other evidence types from receiving zero weight.

3. Equal-source estimand: The equal-source estimand first calculated the within-source status distribution and then averaged those distributions across sources. Equal-source confidence normalized W within each Study_ID before averaging. Composite components are reported separately because aggregation choices can affect interpretation.[26–27]

4. Equal-source confidence estimand: Confidence weights were normalized within each Study_ID before averaging. Because quality attributes were constant within ecological sources, this estimand was numerically identical to the unweighted equal-source estimand.

Uncertainty for source-balanced estimands was quantified using 20,000 cluster-bootstrap samples, where Study_ID values were sampled with replacement and their within-source profiles averaged. These intervals describe variation among documented sources rather than biological uncertainty about unobserved regional stocks.

Spatial verification

Record_ID linked ecological classifications to the repaired D5 crosswalk and Spatial_ID linked D5 to D3. Spatial analyses required Crosswalk_Row_Type = evidence-to-spatial crosswalk, Link_Status = linked, Ecological_Classification_Record = yes and Use_in_Site_Level_Analysis = yes. Coordinate type and validation status remained separate from ecological confidence. Spatial source depth was the number of unique Study_ID values linked to a Spatial_ID. Water bodies were assigned to eight display categories from curated names: river; reservoir or dam; lake; lagoon, estuary or coastal water; wetland, floodplain or delta; stream, creek or inlet; basin or regional context; and other inland water. These categories were used only for visualization and did not alter D3 labels.

Multiplicity simulation

The simulation tested a structural consequence of clustered extraction. The generating population was the empirical equal-source distribution across the four ecological categories. For each replicate, 27 source statuses were sampled from this distribution. Baseline records per source were sampled with replacement from the empirical ecological source-count distribution. Multiplicity for fully exploited sources was then multiplied by ratios from 1 to 10, while source statuses and the number of sources were unchanged. Record-weighted and equal-source distributions were compared with the generating distribution using total-variation distance, $0.5 \times \Sigma k\ |\text{p-hat}(k) - p(k)|$. We also recorded whether each estimator selected a category other than the generating dominant category. Each ratio used 30,000 replicates and a fixed random seed. The simulation separates row-multiplicity bias from source-sampling uncertainty; it is not intended to reproduce all publication and extraction processes.

Software and reproducibility

Analyses were performed in Python using pandas, NumPy, SciPy, statsmodels, GeoPandas and Matplotlib. Country polygons in Fig. 4 were drawn from Natural Earth public-domain data for cartographic context; analytical coordinates were those in D3. The submission package includes source data for every figure, the analysis script, the reference-verification audit and the cleaned D0–D5 tables. Exact package versions and a persistent repository DOI should be inserted before submission.

Table 1 | Analytical units and inferential targets

Unit or estimand	Construction	Question answered	Primary limitation
Record	Each classified D1 row receives equal weight	How is the released database populated?	Dominated by publications contributing many rows
Confidence-weighted record	Rows weighted by indicator breadth, study quality and risk control	How does coded support redistribute database rows?	Repeated source-level attributes may not distinguish records
Equal source	Within-source status profiles averaged across Study_ID values	How do independent sources distribute reported classifications?	All sources receive equal total influence
Equal-source confidence	Confidence weights normalized within Study_ID before averaging	Does within-source support alter the source estimand?	Requires record-specific quality variation
Spatial location	Unique linked Spatial_ID with source-depth count	Where is evidence located and how independently replicated is it?	Point/proxy locations do not represent hydrographic extent

Data availability

The data and analytical materials supporting the findings of this study are available in Zenodo at https://doi.org/10.5281/zenodo.23167024. The repository contains datasets used for reference verification, figure generation, stock-status analyses, spatial verification, simulations, and supporting analytical materials.

References

1. Sayer, C. A. et al. One-quarter of freshwater fauna threatened with extinction. Nature 638, 138–145 (2025). https://doi.org/10.1038/s41586-024-08375-z

2. Reid, A. J. et al. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev. 94, 849–873 (2019). https://doi.org/10.1111/brv.12480

3. Tickner, D. et al. Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. BioScience 70, 330–342 (2020). https://doi.org/10.1093/biosci/biaa002

4. Basurto, X. et al. Illuminating the multidimensional contributions of small-scale fisheries. Nature 637, 875–884 (2025). https://doi.org/10.1038/s41586-024-08448-z

5. Arthur, R. (ed.) Review of the state of world fishery resources—Inland fisheries, fourth edition. FAO Fish. Aquac. Circ. No. 942, Rev. 4 (FAO, Rome, 2024). https://doi.org/10.4060/cd2169en

6. Funge-Smith, S. & Bennett, A. A fresh look at inland fisheries and their role in food security and livelihoods. Fish Fish. 20, 1176–1195 (2019). https://doi.org/10.1111/faf.12403

7. Ainsworth, R., Cowx, I. G. & Funge-Smith, S. J. A review of major river basins and large lakes relevant to inland fisheries. FAO Fish. Aquac. Circ. No. 1170 (FAO, Rome, 2021). https://doi.org/10.4060/cb2827en

8. Lorenzen, K. et al. Stock assessment in inland fisheries: a foundation for sustainable use and conservation. Rev. Fish Biol. Fish. 26, 405–440 (2016). https://doi.org/10.1007/s11160-016-9435-0

9. Abobi, S. M. & Wolff, M. West African reservoirs and their fisheries: an assessment of harvest potential. Ecohydrol. Hydrobiol. 20, 183–195 (2020). https://doi.org/10.1016/j.ecohyd.2019.11.004

10. Amponsah, S. Investigating the fishery of Owabi fishing community in Ghana, West Africa, using traditional ecological knowledge. Int. J. Soc. Sci. Educ. Res. 9, 536–545 (2023). https://doi.org/10.24289/ijsser.1284817

11. Seth, K. et al. Assessing the exploitation status of main fisheries resources in Ghana's reservoirs based on reconstructed catches and a length-based bootstrapping stock assessment method. Lake Reserv. Manag. 35, 415–434 (2019). https://doi.org/10.1080/10402381.2019.1616340

12. Kantoussan, J. et al. Catch per unit effort and yields as indicators of exploited fish communities: application to two West African reservoirs. Lakes Reserv. Res. Manag. 19, 86–97 (2014). https://doi.org/10.1111/lre.12061

13. Elliott, A. M. et al. Modelling catch per unit effort of the Black Volta near the Bui Dam in Ghana. Afr. J. Ecol. 54, 463–470 (2016). https://doi.org/10.1111/aje.12330

14. Sidi, I. I. et al. Demographic parameters and exploitation rate of five key fishes of Okpara Stream, Oueme River, Benin, West Africa. Int. J. Aquat. Biol. 7, 353–361 (2019). https://doi.org/10.22034/ijab.v7i6.642

15. Etim, L., Lebo, P. E. & King, R. P. The dynamics of an exploited population of a siluroid catfish (Schilbe intermedius Rüppell, 1832) in the Cross River, Nigeria. Fish. Res. 40, 295–307 (1999). https://doi.org/10.1016/S0165-7836(98)00217-3

16. Harford, W. J., Babcock, E. A., Carcamo, R. & McDonald, G. A global review of fisheries management and assessment approaches for data-limited fisheries. Front. Mar. Sci. 8, 757877 (2021). https://doi.org/10.3389/fmars.2021.757877

17. Pons, M., Cope, J. M. & Kell, L. T. Comparing performance of catch-based and length-based stock assessment methods in data-limited fisheries. Can. J. Fish. Aquat. Sci. 77, 1026–1037 (2020). https://doi.org/10.1139/cjfas-2019-0276

18. Kokkalis, A. et al. Estimating uncertainty of data-limited stock assessments. ICES J. Mar. Sci. 74, 69–77 (2017). https://doi.org/10.1093/icesjms/fsw145

19. Maunder, M. N. & Punt, A. E. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70, 141–159 (2004). https://doi.org/10.1016/j.fishres.2004.08.002

20. Froese, R. et al. A new approach for estimating stock status from length frequency data. ICES J. Mar. Sci. 75, 2004–2015 (2018). https://doi.org/10.1093/icesjms/fsy078

21. Hurlbert, S. H. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54, 187–211 (1984). https://doi.org/10.2307/1942661

22. Hedges, L. V., Tipton, E. & Johnson, M. C. Robust variance estimation in meta-regression with dependent effect size estimates. Res. Synth. Methods 1, 39–65 (2010). https://doi.org/10.1002/jrsm.5

23. Konstantopoulos, S. Fixed effects and variance components estimation in three-level meta-analysis. Res. Synth. Methods 2, 61–76 (2011). https://doi.org/10.1002/jrsm.35

24. Francis, R. I. C. C. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68, 1124–1138 (2011). https://doi.org/10.1139/f2011-025

25. Thorson, J. T. et al. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. 192, 84–93 (2017). https://doi.org/10.1016/j.fishres.2016.06.005

26. Munda, G. & Nardo, M. Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. Appl. Econ. 41, 1513–1523 (2009). https://doi.org/10.1080/00036840601019364

27. OECD & European Commission Joint Research Centre. Handbook on constructing composite indicators: methodology and user guide (OECD Publishing, Paris, 2008). https://doi.org/10.1787/9789264043466-en

28. Carstensen, J., Murray, C. J. & Lindegarth, M. Getting the errors right: the importance of partitioning sources of uncertainty for ecological indicators. Ecol. Indic. 167, 112637 (2024). https://doi.org/10.1016/j.ecolind.2024.112637

29. Dale, V. H. & Beyeler, S. C. Challenges in the development and use of ecological indicators. Ecol. Indic. 1, 3–10 (2001). https://doi.org/10.1016/S1470-160X(01)00003-6

30. Fulton, E. A., Smith, A. D. M. & Punt, A. E. Which ecological indicators can robustly detect effects of fishing? ICES J. Mar. Sci. 62, 540–551 (2005). https://doi.org/10.1016/j.icesjms.2004.12.012

31. Fischer, S. H. et al. An approach to evaluating and selecting indicators of ecosystem overfishing. Fish Fish. 24, 231–247 (2023). https://doi.org/10.1111/faf.12722

32. Chapman, A. D. & Wieczorek, J. R. (eds.) Georeferencing best practices (GBIF Secretariat, Copenhagen, 2020). https://doi.org/10.15468/doc-gg7h-s853

33. Wieczorek, J., Guo, Q. & Hijmans, R. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. Int. J. Geogr. Inf. Sci. 18, 745–767 (2004). https://doi.org/10.1080/13658810412331280211

34. Darwall, W. R. T. et al. Implications of bias in conservation research and investment for freshwater species. Conserv. Lett. 4, 474–482 (2011). https://doi.org/10.1111/j.1755-263X.2011.00202.x

35. Achieng, A. O. et al. Monitoring biodiversity loss in rapidly changing Afrotropical ecosystems: an emerging imperative for governance and research. Philos. Trans. R. Soc. B 378, 20220271 (2023). https://doi.org/10.1098/rstb.2022.0271

36. Masese, F. O., Arimoro, F. O., Dalu, T. & Gettel, G. M. Editorial: freshwater science in Africa. Front. Environ. Sci. 11, 1233932 (2023). https://doi.org/10.3389/fenvs.2023.1233932

37. Affinito, F. et al. Assessing coverage of the monitoring framework of the Kunming–Montreal Global Biodiversity Framework and opportunities to fill gaps. Nat. Ecol. Evol. 9, 1280–1294 (2025). https://doi.org/10.1038/s41559-025-02718-3

38. Smits, A. P. Too much and not enough data: challenges and solutions for generating information in freshwater research and monitoring. Ecosphere 16, e70205 (2025). https://doi.org/10.1002/ecs2.70205

39. van Rees, C. B. et al. Grasping at water: a gap-oriented approach to bridging shortfalls in freshwater biodiversity conservation. Biol. Rev. 100, 1970–1993 (2025). https://doi.org/10.1111/brv.70030

40. Nicholson Thomas, I. et al. Towards a standardised monitoring of ecosystem condition: a literature review on indicators and their data sources. Ecol. Indic. 178, 113952 (2025). https://doi.org/10.1016/j.ecolind.2025.113952

41. Ouédraogo, R. B., Sanogo, S. & Compaoré, I. Population dynamics of Nile tilapia, Oreochromis niloticus, in Samendéni Reservoir, Burkina Faso. Fish. Aquat. Sci. 28, e25 (2025). https://doi.org/10.47853/FAS.2025.e25

42. Ogueri, C. et al. Stock status of Hepsetus odoe in River Imo, Nigeria, for conservation and management strategies for sustainability. Manglar 22, 285–295 (2025). https://doi.org/10.57188/manglar.2025.031

43. Nomwine, N. et al. Growth and stock status assessment of African catfish, Clarias anguillaris, from Burkina Faso newly man-made Lake Samandeni. Adv. Biosci. Biotechnol. 14, 339–356 (2023). https://doi.org/10.4236/abb.2023.147022

44. Samuel, A. et al. Stock assessment indicators for sustainable exploitation of Chrysichthys walkeri in Lake Volta, Ghana, West Africa. J. Limnol. Freshw. Fish. Res. 8, 155–168 (2022). https://doi.org/10.17216/limnofish.1037218

45. Souley, M. et al. Exploitation parameters for the Mormyrus rume population in the River Niger in Niamey, Niger, West Africa. Egypt. J. Aquat. Biol. Fish. 28, 1489–1503 (2024). https://doi.org/10.21608/ejabf.2024.389789

46. Momodou, J. et al. Assessment of growth and stock status of data-limited Ethmalosa fimbriata fishery in Gambia. Reg. Stud. Mar. Sci. 79, 103786 (2024). https://doi.org/10.1016/j.rsma.2024.103786

47. Meulenbroek, P. et al. Fish communities, habitat use, and human pressures in the Upper Volta Basin, Burkina Faso, West Africa. Sustainability 11, 5444 (2019). https://doi.org/10.3390/su11195444

48. Laë, R. Climatic and anthropogenic effects on fish diversity and fish yields in the Central Delta of the Niger River. Aquat. Living Resour. 8, 43–58 (1995). https://doi.org/10.1051/alr:1995004

49. Mensah, E. T. D. et al. Seasonal changes in fish catch and environmental variables in a large tropical lake, Volta, Ghana. Afr. J. Ecol. 57, 66–75 (2019). https://doi.org/10.1111/aje.12537

50. Oluwadamilare, K. et al. Assessment of climatic and environmental parameters on fish abundance of an Afro-tropical reservoir. Sci. Rep. 14, 23561 (2024). https://doi.org/10.1038/s41598-024-75181-y

51. Simier, M., Écoutin, J.-M. & Tito de Morais, L. The PPEAO experimental fishing dataset: fish from West African estuaries, lagoons and reservoirs. Biodivers. Data J. 7, e31374 (2019). https://doi.org/10.3897/BDJ.7.e31374

52. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71 (2021). https://doi.org/10.1136/bmj.n71

53. Haddaway, N. R., Macura, B., Whaley, P. & Pullin, A. S. ROSES reporting standards for systematic evidence syntheses. Environ. Evid. 7, 7 (2018). https://doi.org/10.1186/s13750-018-0121-7

54. Ed-Idoko, J. O., Kindong, R., Tian, S. & Osei, S. A. Dataset supporting: *Regional stock status inference shaped by source dependence in inland fisheries* [Dataset]. Zenodo (2026). https://doi.org/10.5281/zenodo.23167024

Acknowledgements

Authors acknowledge the use of AI-assisted tools for language editing and manuscript drafting. Authors reviewed, validated, and take full responsibility for all scientific content, intellectual content, data interpretation, and conclusions.

Funding

This research received no external funding.

Author Contributions

J.O.E.: conceptualization, data curation, investigation, methodology, visualization and writing original draft.

R.K.: methodology, supervision, validation, review and correspondence.

T.S.: supervision, project administration, methodology and review.

S.A.O.: review

Competing Interests

The authors declare no competing interests.

Supplementary information

Supplementary Fig. 1 | Semantic audit of Confidence_Level and country-level descriptive profiles. a, Cross-tabulation of Confidence_Level and Bias_Risk codes. b, Source-balanced adverse classification shares by country; labels report the number of independent Study_ID–country clusters. Country values supported by one or two sources are discovery indicators rather than national estimates.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- graphicalabstract.png