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.

Participation, Persistence and Influence in an Expanding Research Field: Scientific Workforce Structure in Bentonite–Cementitious Research, 1981–2026

Ayokunle Odunayo Alade
Universiti Sains Malaysia (USM)
Shoaib Hassan
Universiti Sains Malaysia (USM)
Mineesha Sivakumar
Universiti Sains Malaysia (USM)
Shangeetha Ganesan
Dream Catcher Academy Sdn Bhd
Mazidatulakmam Miskam
mazidatul@usm.my

Universiti Sains Malaysia (USM)

Research Article

Keywords: research evaluation methodology, scientific workforce, longitudinal recurrence, scholarly influence, author name disambiguation, bibliometrics

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11091332/v1

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

Additional Declarations: No competing interests reported.

Abstract

Standard bibliometric indicators of research-field growth, such as contributor counts and publication totals, are widely used in research evaluation despite limited direct testing of whether they track recurring participation and scholarly influence rather than one-time entry. We test this distinction using bentonite–cementitious research, an interdisciplinary materials-science domain in which annual output rose from 23 documents in 2012 to 179 in 2025, analysing 1,492 English-language Web of Science Core Collection articles and reviews (1981–2026) through author productivity, full and fractionalised publication counts, Lotka-law fits estimated by ordinary least squares and maximum likelihood, institutional concentration, longitudinal affiliation trajectories, and citation influence. Of 4,752 bibliographic author records, 78.28% appeared in one publication. Productivity concentration exceeded the classical Lotka exponent under ordinary least squares ($\beta = 2.54$) and discrete maximum-likelihood estimation ($\alpha = 2.72$, 95% confidence interval [2.68, 2.77]); an audited sensitivity check did not reduce concentration among the leading records examined. Institutional output was comparatively dispersed (Gini = 0.39; HHI = 14/10,000), with the leading 15 of 1,674 affiliations accounting for 8.0% of distinct-document output. Recurrence was more restricted: 72.2% of affiliations published in one year and 77.4% within one five-year period. Full and fractionalised counts produced different productivity profiles, while publication frequency and citation influence identified partly overlapping contributor groups. The Participation–Persistence–Influence framework therefore treats these as partially coupled, level-dependent dimensions that should be reported separately while disclosing author-identity uncertainty.

1. Introduction

Bibliometric indicators, counts of publications, contributors, or citations, are widely used as evidence of a research field's development and of individual or institutional research performance (Donthu et al., 2021; Waltman & van Eck, 2015). Their use rests on an assumption that is rarely tested directly: that growth in these counts corresponds to a correspondingly broad, recurrent and influential underlying research workforce. Whether contributor counts, productivity measures and citation-based indicators in fact identify the same population, or instead measure related but non-equivalent properties of scientific activity, is therefore a methodological question with direct consequences for how research fields and their contributors are evaluated.

The organisation of the scientific workforce underlying such interdisciplinary research is also changing. Scientific knowledge production has increasingly shifted towards team-based research, particularly where complex problems require expertise distributed across disciplinary and institutional boundaries (Wu et al., 2019). Longitudinal studies of scientific careers further show that growth in the number of researchers entering scientific domains does not necessarily translate into equally sustained participation. Scientific workforces may include large populations of temporary or occasional contributors alongside substantially smaller groups that publish recurrently over longer periods (Gordon, 2007; Milojević et al., 2018). Consequently, growth in publications or contributor numbers alone provides limited evidence about the continuity of the knowledge-producing workforce underlying a research field.

This question is examined here using bentonite–cementitious research as a case. The field sits at the interface of cement chemistry, clay mineralogy, geotechnical engineering and alternative-binder research, and its expansion has been driven in part by the wider transition towards lower-carbon construction, which places increasing emphasis on supplementary cementitious materials, alternative binders and calcined-clay systems (IEA, 2025; GCCA, 2025). Existing reviews of this literature have primarily addressed material behaviour and engineering performance, including the current status and future prospects of bentonite in cement-based composites (Hasan et al., 2024); they do not examine how participation in the underlying contributor population is distributed, distinguish participation, recurrence and scholarly influence as separate properties, or test whether standard growth indicators track these properties. The field's rapid, interdisciplinary growth, combined with the absence of prior workforce-structure analysis, makes it a suitable case for the methodological question set out above.

The distinction is important because participation, recurrence and scholarly influence represent related but non-equivalent properties of scientific activity. Earlier authorship research distinguishes transient contributors from researchers who publish recurrently within a field (Gordon, 2007), while longitudinal studies demonstrate substantial heterogeneity in the duration and intensity of scientific publication careers (Petersen et al., 2012; Milojević et al., 2018). Research on cumulative advantage further indicates that scholarly recognition is unevenly distributed and cannot be inferred directly from publication volume (Merton, 1968; Allison et al., 1982). These literatures establish that participation, continuity and recognition are individually important, but they do not establish that the three dimensions necessarily identify the same structure within a specific interdisciplinary research domain.

This creates an analytical problem for field-level bibliometric assessment. A research domain may contain a large number of contributors because many researchers participate occasionally, while recurrent publication remains concentrated within a much smaller population. Similarly, researchers who publish most frequently may not be those whose work receives the greatest scholarly recognition. Treating contributor counts, publication productivity or citations as interchangeable indicators of scientific leadership can therefore obscure differences between entry into a field, recurrent contribution and accumulated scholarly influence.

The present study therefore distinguishes three dimensions of scientific workforce structure. Participation concerns the breadth and distribution of contribution to the research domain. Persistence is the conceptual longitudinal dimension and is operationalised empirically as observable recurrence of bibliographic participation across publication years, rather than uninterrupted career activity. Influence concerns scholarly visibility represented through domain-specific and broader citation indicators. The three dimensions may converge for some contributors but are not assumed to be interchangeable or causally ordered.

The study is deliberately restricted to author- and affiliation-level workforce structure. It does not reconstruct citation networks, conceptual themes, journal structures, national productivity or

geographical collaboration. Accordingly, the unit of analysis remains centred on the bibliographically observable knowledge-producing workforce.

The study addresses four research questions:

RQ1. How is scientific participation distributed among authors in bentonite–cementitious research?

RQ2. To what extent do productivity and scholarly influence identify the same contributors?

RQ3. What longitudinal patterns of recurrent institutional participation are evident in this research domain?

RQ4. What does the relationship among participation, longitudinal recurrence and influence reveal about the structure of the research workforce?

To address these questions, the study develops the Participation–Persistence–Influence (P–P–I) framework as an analytical device rather than a causal theory or performance index. Its purpose is to examine whether growth in contributor numbers is accompanied by similarly distributed recurrent participation and scholarly influence. The contribution therefore lies not in proposing another ranking of researchers, but in distinguishing field size, recurrence of participation and scholarly visibility as separate dimensions of bibliographically observable workforce structure. More broadly, the analytical strategy applied here, testing multiple counting conventions, comparing independent estimators, and auditing bibliographic-identity assumptions against the concentration statistics they support, offers a transferable approach for evaluating the validity of participation, productivity and influence indicators in other bibliometrically observable research fields, beyond the domain examined here.

2. Materials and Methods

2.1 Research design

A longitudinal bibliometric performance-analysis design was used to examine the scientific workforce structure of bentonite–cementitious research (Donthu et al., 2021). Three analytical dimensions were distinguished: participation, persistence and scholarly influence.

Participation was examined through authorship structure, author-productivity distributions, and full and fractionalised publication counts. Persistence was operationalised as longitudinal bibliographic recurrence and assessed through author- and affiliation-production trajectories. Influence was examined using local citations, total citations, h-index, g-index and m-index measures.

The dimensions were retained separately because they represent different properties of the publication workforce. No composite leadership, productivity or influence score was constructed, and no causal progression among participation, persistence and influence was assumed.

The unit of analysis was restricted to author and affiliation records. Document-, author- and source-level co-citation analysis, historiographic analysis, reference publication year spectroscopy, keyword co-occurrence, thematic mapping, thematic evolution, journal productivity, country productivity and international collaboration were outside the analytical scope.

2.2 Data source and search strategy

Bibliographic records were retrieved from the Web of Science Core Collection (WoSCC) on 1 August 2026 using the Topic Search expression:

TS=(bentonite AND (cement* OR geopolymer* OR "alkali-activated" OR "alkali activated"))

The search required bentonite to occur with terminology representing conventional cementitious or alternative-binder systems. The wildcard cement* accommodated lexical variants associated with cement and cementitious materials, while the additional terms extended retrieval to geopolymer and alkali-activated research.

The search boundary was intentionally application-inclusive. It was not restricted to studies in which bentonite was used directly as a Portland-cement replacement because this would exclude parts of the scientific workforce working on cement–clay interactions, geotechnical applications and alternative binder systems.

The corpus covered 1981–2026. Because retrieval occurred on 1 August 2026, the final year represents a partial observation period and was interpreted cautiously in longitudinal analysis. Records were indexed in the Science Citation Index Expanded (SCI-EXPANDED; n = 1,248), the Emerging Sources Citation Index (ESCI; n = 222), SCI-EXPANDED jointly with the Social Sciences Citation Index (SSCI; n = 21), and SCI-EXPANDED jointly with the Arts & Humanities Citation Index (A&HCI; n = 1).

2.3 Corpus construction

The WoSCC Topic Search produced 1,880 records. Restriction to the WoS Article and Review document types reduced the dataset to 1,529 records; this restriction followed standard bibliometric practice of excluding conference proceedings papers, editorials and other non-standard formats because their citation and indexing behaviour is not directly comparable to that of journal articles and reviews. Review articles were retained because they represent a normal and often influential contribution format in this applied materials-science field, though the restriction means the corpus may under-represent contributors whose participation in this domain is weighted toward conference proceedings, a possibility we return to in the Limitations. Subsequent restriction to English-language publications produced a final analytical corpus of 1,492 articles and reviews (1,444 articles and 48 reviews).

The screening sequence was therefore:

1,880 retrieved records → 1,529 Article/Review records → 1,492 English-language articles/reviews → 1,492 articles and reviews analysed (Fig. 1).

Note

Adapted from the PRISMA flow-diagram format for a single-database bibliometric corpus (Web of Science Core Collection); this is not a formal PRISMA 2020 systematic-review diagram.

No additional topical exclusion was undertaken after construction of the final corpus. Because all records originated from a single WoSCC retrieval rather than from merged bibliographic databases, cross-database deduplication was not required.

2.4 Data processing and bibliographic identities

Bibliographic processing was conducted using R 4.5.1 and Bibliometrix/Biblioshiny 5.4.1, with RStudio Desktop 2026.07.1–147 as the analytical environment (Aria & Cuccurullo, 2017). The analysis used author names, affiliations, publication years, authorship information and citation metadata extracted from the final WoSCC corpus.

Author productivity was evaluated using both full and fractionalised counting. Under full counting, each co-author received one publication count for a document. Under fractionalised counting, publication credit was divided according to the number of authors associated with the document. The two measures were retained separately because they represent different aspects of participation and were not combined into a composite productivity measure (Waltman & van Eck, 2015).

Longitudinal author and affiliation outputs were generated from publication-year information in the processed bibliographic records. The trajectories were used descriptively to identify recurrent bibliographic presence and differences in the timing and intensity of contribution. They were not used to reconstruct complete scientific careers.

Author-level results require particular caution because WoSCC metadata can contain abbreviated names, name variants and homonyms, so common surname–initial combinations may represent more than one individual (Strotmann & Zhao, 2012). Author labels are therefore treated as bibliographic records rather than verified individual identities unless separately disambiguated; the same caveat applies to affiliations, where recurrence indicates bibliographic presence rather than continuity of the same researchers or programme.

2.5 Scientific participation

RQ1 was examined at author level using corpus-level authorship indicators and the distribution of publication frequency.

Participation was characterised using the number of author records, single-authored documents, authors associated with single-authored publications and mean co-authors per document.

The distribution of author productivity was examined using publication-frequency data and the Lotka output generated by Biblioshiny, following Lotka's (1926) distributional model of scientific productivity.

The fitted exponent was estimated by ordinary least-squares regression of log10(number of authors) on log10(number of documents) across the 20 non-empty productivity classes in the Biblioshiny Lotka table, following the standard log-log linearisation of Lotka's law; a Kolmogorov–Smirnov test (Pao, 1985) compared the empirical distribution against the classical theoretical exponent of 2.0. As a robustness check reported alongside the OLS estimate (Section 3.1), the exponent was additionally estimated by discrete maximum-likelihood estimation (MLE) following Clauset, Shalizi and Newman (2009), which is more statistically efficient than OLS for this class of distribution; automatic lower-bound (xmin) selection and a log-likelihood-ratio comparison against lognormal and exponential alternatives were also performed.

Author productivity was further evaluated using full and fractionalised publication counts. Full counting represented frequency of publication participation, whereas fractionalised counting represented authorship-adjusted contribution.

2.6 Scholarly influence

RQ2 examined whether publication productivity and scholarly influence identified similar contributor profiles.

Influence measures comprised local citations, total citations, h-index, g-index and m-index. Local citations represent citations received from documents within the 1,492-article analytical corpus and were therefore interpreted as domain-specific scholarly visibility. Total citations represent broader citation visibility within the coverage of the bibliographic database.

The h-index represents the relationship between publication output and accumulated citations (Hirsch, 2005), while the g-index gives greater weight to highly cited publications (Egghe, 2006). The m-index relates the h-index to the duration of the indexed publication history and was therefore used as a time-normalised indicator.

The m-index was not interpreted as a measure of actual career duration because the first publication represented within the retrieved corpus may not coincide with the beginning of an author's scientific career. None of the citation indicators was interpreted as a direct measure of scientific quality.

No composite influence score was created.

2.7 Persistence as longitudinal bibliographic recurrence

RQ3 examined persistence through affiliation-production-over-time outputs. Author-level longitudinal recurrence was assessed during analysis but is not reported at individual level, for the reasons given in Section 4.6 (Limitations).

Persistence constitutes the conceptual dimension of the framework; longitudinal bibliographic recurrence constitutes its empirical operationalisation. Repeated appearance of an author record across multiple publication years was interpreted as evidence of recurrence within the focal literature. No

minimum number of active years, continuity ratio or formal persistence threshold was imposed. For every affiliation with document-level presence in the corpus, persistence was additionally characterised using the number of distinct active publication years, the number of active 5-year periods (Section 3.3; Table 4), the span between first and last indexed year, the proportion of that span spent active, and the longest publication-free gap between active years; affiliations were classified as one-off, consecutive, short-gap recurrent or intermittent on this basis (Online Resource 2). Institutional output was counted on a distinct-document basis, so an affiliation appearing on a document was counted once regardless of how many co-authors on that document shared the affiliation. Institutional output concentration was assessed using the Gini coefficient and the Herfindahl–Hirschman Index (HHI) alongside top-5/10/15 output shares, computed over the full population of affiliations represented in the corpus rather than the leading institutions alone.

Long-duration recurrence was distinguished descriptively from more recently concentrated recurrence. These categories refer solely to temporal publication patterns and do not indicate uninterrupted research activity, employment continuity, career survival or researcher commitment.

At affiliation level, recurrent publication under the same institutional identity was used to characterise organisational recurrence. This does not imply continuity of particular researchers or research groups.

2.8 Participation–Persistence–Influence synthesis

RQ4 was addressed through structured comparison of the evidence generated for RQ1–RQ3 rather than through an additional statistical model.

Participation was represented by authorship distribution and productivity; persistence was empirically represented by longitudinal bibliographic recurrence; and influence was represented by citation and impact indicators.

The framework does not assume a causal sequence of participation → persistence → influence.

Instead, it evaluates whether the three dimensions converge or diverge within the recorded workforce.

The framework is therefore interpretive and descriptive rather than predictive or causal.

2.9 Research question–method alignment

Table 1 summarises how each research question maps to its unit of analysis, analytical procedure and intended interpretation.

Table 1
Alignment of research questions, analytical units, procedures and interpretation

Research question	Unit of analysis	Analytical procedure	Principal output	Intended interpretation
RQ1. Distribution of participation	Authors	Authorship indicators; productivity distribution; full and fractional counts	Workforce and productivity profiles	Breadth and concentration of participation
RQ2. Productivity versus influence	Authors	Local/total citations; h-, g- and m-indices	Productivity and impact profiles	Convergence/divergence between output and influence
RQ3. Longitudinal recurrence	Authors; affiliations	Author and affiliation production over time	Longitudinal trajectories	Descriptive recurrence of contribution
RQ4. Workforce configuration	Authors; affiliations	Structured synthesis of RQ1–RQ3	P–P–I framework	Relationship among participation, recurrence and influence
Note: Participation, persistence and influence are analytically distinct. RQ4 does not constitute a composite index, causal model or statistical validation of the framework.				

3. Results

Annual output grew over the observation window: 188 documents were published before 2010, 392 in 2010–2019 and 912 in 2020–2026, with annual counts rising from 23 in 2012 to a peak of 179 in 2025 (2026 is a partial year; Section 2.2). This growth motivates, but is analytically separate from, the participation, persistence and influence distinctions reported below.

3.1 Broad participation and concentration of recurrent contribution

The corpus contained 4,752 author records across 1,492 documents, with an average of 4.75 co-authors per document. Scientific participation was highly unevenly distributed.

Of the 4,752 author records, 3,720 (78.28%) appeared in one publication, and a further 590 (12.42%) appeared in two. Thus, 90.70% of the recorded author population contributed no more than two documents.

Table 2
Distribution of author participation by publication frequency

Documents	Author records	Share (%)
1	3,720	78.28
2	590	12.42
3	201	4.23
4	87	1.83
5	59	1.24
6–10	66	1.39
> 10	29	0.61
Total	4,752	100.00

The distribution reveals a marked separation between broad participation and recurrent contribution (Fig. 2). Although several thousand author records entered the literature, repeated publication was concentrated within a substantially smaller proportion of the contributor population.

A formal test of the productivity distribution supports this description. Biblioshiny's Lotka fit (20 productivity classes, from 1 to 27 documents) estimated $\beta = 2.54$ ($C = 0.5841$, $R^2 = 0.9446$), steeper than the classical theoretical Lotka exponent of 2.0. A Kolmogorov–Smirnov test against the classical $\beta = 2.0$ distribution was significant ($p = .035$), indicating that productivity in this domain is more concentrated among a small set of author records than the classical approximation predicts; a second Kolmogorov–Smirnov test against the domain-specific fitted exponent ($\beta = 2.54$) was not significant ($p = .978$), confirming that the fitted model, though more skewed than the textbook case, describes the empirical distribution closely.

To test whether this concentration estimate is inflated by the author-identity limitation discussed in Section 4.6, the exponent was refitted after substituting audited individual-level counts (Online Resource 1) for leading-author records confirmed to be composites of more than one real individual. Of the thirteen leading-author records audited (Section 4.6), three (Cuevas J, Chen Y, Plecas I) were confirmed to be essentially single individuals and were left at their naive count; the remaining ten were confirmed composite. Nine of these ten composite records could be individually corrected; the tenth (Li Z, productivity class $x = 13$) shares its productivity class with seven further, non-leading-author records that were not individually audited, so the class as a whole was left at its naive value rather than partially corrected. The productivity class containing Plecas I ($x = 15$) was left unchanged for the same reason. The corrected distribution (17 productivity classes) yielded $\beta = 2.82$ ($R^2 = 0.968$), higher than the uncorrected $\beta = 2.54$, and remained significantly more concentrated than the classical exponent (Kolmogorov–Smirnov $D = 0.47$, $p = .045$) while closely fitting its own exponent ($D = 0.12$, $p > .99$).

Correcting for the disambiguation limitation among the thirteen audited leading records therefore did not reduce the estimated concentration. Because only those thirteen records were audited, this check does not establish that the reported exponent is unaffected by identity errors elsewhere in the 4,752-record population; it establishes only that the disambiguation limitation does not explain away the concentration observed among the records actually checked.

As a robustness check on the estimation method itself, the exponent was re-estimated by discrete maximum-likelihood estimation (MLE; Clauset et al., 2009), which is more statistically efficient than OLS for this class of distribution. The naive productivity distribution yielded α = 2.72 (95% CI [2.68, 2.77]; asymptotic and bootstrap intervals coincided), and the disambiguation-corrected distribution yielded α = 2.74 (95% CI [2.69, 2.79]); automatic lower-bound selection confirmed xmin = 1 as the best-fitting threshold, so no part of the productivity range was excluded from the estimate. Both MLE estimates exceed the classical exponent of 2.0 by a wide margin, corroborating the OLS-based conclusion above under a more efficient estimator. A log-likelihood-ratio comparison nonetheless found that a lognormal distribution fits the naive productivity data significantly better than a pure power law ($R = -4.94$, $p = .034$), while a power law clearly outperforms an exponential null ($R = 491.6$, $p < .001$). Lotka's law is conventionally expressed as a power law, and this manuscript retains that convention for comparability with the bibliometric literature, but the lognormal comparison indicates that the power-law form should not be read as the uniquely correct description of this productivity distribution.

3.2 Productivity, counting method and scholarly influence: population-level associations

Author-level productivity and influence were compared using population-level statistics rather than named leading-author rankings, for the reasons explained in Section 4.6.

The treatment of shared authorship materially affects the productivity ranking: full-count and fractionalised productivity correlate only moderately across the full author population (Spearman's $\rho = 0.61$, $n = 4{,}752$, $p < .001$), confirming quantitatively that counting method reshapes the apparent distribution of productivity.

Productivity and scholarly influence identify overlapping but non-equivalent contributor profiles. Full-count productivity correlates only moderately with local citation counts ($\rho = 0.32$, $n = 4{,}752$, $p < .001$), the domain-specific influence measure, and with total citations ($\rho = 0.40$, $n = 4{,}289$, $p < .001$). The correlation between full-count productivity and h-index is considerably stronger ($\rho = 0.91$, $n = 4{,}289$; $n = 4{,}289$ reflects the subset of author records for which Biblioshiny's impact-indicator table reports a citation-based index, a smaller population than the 4,752 records used for the productivity and local-citation correlations), but this should be read cautiously: h-index is mathematically bounded by publication count, so a strong correlation between the two is expected on structural grounds rather than constituting independent evidence that productivity and influence converge. The same pattern holds even more strongly for g-index ($\rho = 0.99$, $n = 4{,}289$), which is mechanically weighted toward productive authors by construction, and does not hold to the same degree for the time-normalised m-index ($\rho = 0.33$, $n = 4{,}289$, $p < .001$),

which correlates only moderately with productivity, consistent with the m-index capturing a genuinely different, less productivity-bound dimension of scholarly recognition.

RQ2 therefore identifies productivity and scholarly influence as overlapping but non-equivalent dimensions at the population level. Because individual author identity cannot currently be verified across common surname-initial combinations (Section 4.6), this evidence is reported only as population-wide association, not as a ranking of named contributors.

3.3 Institutional recurrence

Although institutional publication output was widely distributed, recurrent participation across multiple years was concentrated among a comparatively small subset of affiliations.

Institutional output was considerably less concentrated than author-level productivity. Across the 1,674 distinct affiliations represented in the corpus, distinct-document output produced a Gini coefficient of 0.39 and a Herfindahl–Hirschman Index of 14 (0–10,000 scale); the leading 15 affiliations listed in Table 3 accounted for 8.0% of total institutional output (top-5 = 3.6%; top-10 = 5.9%). What distinguishes recurrent affiliations is therefore not a dominant share of total output but sustained bibliographic presence: 72.2% of the 1,674 affiliations published in only a single year and 77.4% concentrated all of their output within a single 5-year period, whereas the affiliations in Table 3 are drawn disproportionately from the smaller group publishing across multiple years and periods (Online Resource 2 reports the full affiliation-level distribution).

Table 3
Leading institutional affiliations by distinct-document output

Rank	Affiliation	Distinct publications
1	Tongji University	31
2	Southeast University	26
3	China University of Mining and Technology	20
4	Hohai University	19
5	Central South University	16
5	Universidad Autónoma de Madrid	16
5	University of Cambridge	16
8	University of Belgrade	14
8	University of Engineering and Technology	14
10	Chang'an University	13
10	Shandong University	13
10	Southwest University of Science and Technology	13
10	University of Bern	13
10	University of British Columbia	13
10	Vinča Institute of Nuclear Sciences	13

Longitudinal affiliation patterns showed that leading institutions accumulated distinct publications across different observation spans and with different degrees of continuity.

Table 4
Longitudinal recurrence indicators for the five leading institutional affiliations by distinct-document output

Affiliation	First year	Last year	Distinct docs	Active years	5-year periods	Max gap (yr)
Tongji University	2014	2026	31	12	4	1
Southeast University	2013	2026	26	10	4	4
China Univ. Mining & Technol.	2019	2026	20	8	3	0
Hohai University	2012	2026	19	10	4	3
Central South University	2018	2026	16	8	3	1

Table 4 reports longitudinal recurrence indicators for the five leading affiliations identified in Table 3. Three entered the corpus earlier (Hohai University, 2012; Southeast University, 2013; Tongji University, 2014), while Central South University (2018) and China University of Mining and Technology (2019) entered later. Their recurrence patterns nevertheless differ. China University of Mining and Technology was active in all eight years of its observed span, with no publication-free gap; Tongji University was active in 12 of 13 years, with a longest gap of one year; and Southeast University was active in 10 of 14 years, with a longest gap of four years. The leading affiliations therefore combine different durations and degrees of continuity rather than following a single temporal pattern.

3.4 Integrated workforce configuration

The combined evidence indicates that participation, longitudinal recurrence and scholarly influence overlap but do not describe the same workforce structure.

Table 5
Empirical correspondence among participation, longitudinal recurrence and influence

Workforce pattern	Empirical basis	Interpretation
Broad participation, limited recurrence	78.28% contributed once; 90.70% contributed ≤ 2 publications	Entry is substantially more diffuse than repeated contribution
Counting-method sensitivity	Full-count and fractionalised productivity correlate only moderately (ρ = 0.61)	Productivity depends partly on treatment of shared authorship
Productivity–influence association	Productivity correlates only moderately with local citations (ρ = 0.32) and total citations (ρ = 0.40)	Influence is related to, but not reducible to, output volume
Long-duration institutional recurrence	Earlier-entering leading affiliations (Hohai, Southeast and Tongji) span 13–15 observed years and 10–12 active publication years (Table 4)	Long-duration organisational recurrence can include intermittent gaps
Recent concentrated institutional recurrence	China University of Mining and Technology and Central South University entered in 2018–2019 and were active in 8 of 8–9 observed years (Table 4)	More recent recurrence can be highly continuous
Institutional variation in trajectory shape	Longest publication-free gaps among the five leading affiliations range from 0 to 4 years (Table 4)	Organisational recurrence differs in duration and continuity

Note: The patterns summarised in Table 5 **synthesise the descriptive evidence reported in** Sections 3.1–3.3. **They do not constitute a composite index, statistical classification or causal validation of the P–P–I framework. Individual author-level evidence is not included, for the reasons given in** Section 4.6.

The central empirical result is therefore structural rather than rank-based: participation is highly diffuse at entry, distinct-document institutional output is comparatively dispersed and unconcentrated by Herfindahl–Hirschman standards, yet multi-year institutional recurrence is concentrated among a small

minority of affiliations, and scholarly influence follows a partly different distribution from publication frequency. These findings provide the basis for evaluating the Participation–Persistence–Influence framework in the Discussion.

4. Discussion

4.1 Broad participation does not imply a broad recurrent workforce

The clearest structural feature of the bentonite–cementitious research workforce is the difference between the breadth of participation and the much narrower distribution of recurrent contribution. Although 4,752 bibliographic author records were represented in the corpus, 78.28% appeared in only one publication and 90.70% in no more than two. The field therefore combines extensive participation with a comparatively small recurrent contributor population.

This distinction is consistent with earlier work separating transient from continuing contributors in scientific fields. Gordon (2007), for example, showed that research communities may contain large peripheral populations alongside much smaller groups of recurrent authors. Longitudinal studies of scientific careers similarly demonstrate that growth in the number of people entering research does not necessarily produce comparable growth in long-duration participation (Milojević et al., 2018). The present findings extend that distinction to an interdisciplinary materials domain and show why the size of an author population should not be treated as a direct indicator of durable research capacity: what the bibliographic record supports is a narrower claim about observed recurrence, not about underlying research capacity, funding or institutional investment, none of which is directly measurable from publication counts.

This does not make occasional contributors unimportant: interdisciplinary fields often benefit from researchers entering from neighbouring specialties. It does mean entry and recurrence are different structural properties that field-development evaluation should separate (Section 4.5).

This distinction forms the first element of the Participation–Persistence–Influence framework. Participation captures the breadth of contribution, whereas persistence is represented more conservatively here by observable longitudinal recurrence. The results show that the two dimensions do not expand proportionately: a field can attract many contributors while recurrent publication remains concentrated among relatively few.

4.2 Productivity, counting method and scholarly influence form partly separate patterns

The moderate correlation between full-count and fractionalised productivity ($\rho = 0.61$) demonstrates that scientific productivity is partly dependent on how collaborative authorship is treated, rather than being a single, unambiguous property. This matters because team-based research has become increasingly

prominent across science and engineering, particularly in interdisciplinary fields where specialised expertise is distributed among multiple contributors (Wu et al., 2019). Full counting identifies how frequently an author appears in published outputs; fractionalised counting adjusts for shared authorship. Neither measure provides a complete representation on its own, and both were retained separately within the participation dimension of the framework rather than combined into a composite productivity score.

A stronger distinction emerged when productivity was compared with scholarly influence. The moderate correlations between full-count productivity and local citations ($\rho = 0.32$) and total citations ($\rho = 0.40$) indicate that publication frequency and citation-based recognition identify overlapping but far from identical contributor populations. This divergence challenges the common practice of treating prolific authorship as synonymous with scientific leadership. Publication frequency represents contribution to knowledge production; citation influence represents subsequent scholarly recognition. These processes may reinforce each other, but they are not equivalent, consistent with earlier research on cumulative advantage and the uneven distribution of scientific recognition (Merton, 1968; Allison et al., 1982).

The stronger correlation between productivity and h-index ($\rho = 0.91$) requires separate comment, because h-index is mathematically bounded by publication count; the correlation is therefore expected on structural grounds and should not be read as independent evidence that productivity and influence converge. More broadly, accumulated-citation indicators and time-normalised indicators such as the m-index capture different temporal aspects of scholarly recognition, and comparisons based solely on cumulative metrics can favour researchers with longer publication histories (Petersen et al., 2012). This is borne out at the population level: g-index correlates almost perfectly with productivity ($\rho = 0.99$), consistent with its construction, while the m-index correlates only moderately ($\rho = 0.33$), indicating that it captures a meaningfully different, more time-sensitive dimension of scholarly recognition than either h- or g-index. These indices are not reported at individual level in this manuscript (Section 4.6); the population-level pattern is nonetheless informative on its own.

This strengthens the case for treating influence as an independent dimension: the population-level correlations show a real but partial association with productivity, not equivalence.

4.3 Longitudinal recurrence occurs through different temporal configurations

The institutional recurrence indicators (Table 4; Section 3.3) show that repeated participation does not follow one temporal form. Some leading affiliations have longer observation spans with intermittent gaps, whereas others entered more recently but published in most or all years since first appearing in the corpus. Persistence in this framework therefore concerns repeated bibliographic presence across time, not a single pattern of continuous or high-volume output.

Longitudinal studies of scientific careers report comparable heterogeneity in duration, intensity and continuity at the individual level (Petersen et al., 2012; Milojević et al., 2018); the present institutional-

level results extend this observation to organisational participation in bentonite-cementitious research specifically. This distinction also clarifies the use of persistence in the proposed framework: it denotes repeated bibliographic presence across time, not uninterrupted institutional activity or personnel continuity (see Section 2.7), and does not imply continuity of the same researchers or research groups.

Author-level longitudinal recurrence was examined during this study but is not reported at individual level, because the underlying author identities could not be reliably disambiguated across common surname-initial combinations (Section 4.6). At institutional level, however, recurrence can be characterised directly: across the full population of 1,674 affiliations, 72.2% publish in only a single year and 77.4% concentrate their output within a single 5-year period (Section 3.3; Online Resource 2), confirming that the affiliations highlighted in Table 4 are drawn from a small recurrent minority rather than being representative of institutional participation generally. Even within that minority, recurrence is not uniformly continuous: among affiliations active in two or more years, only 21.7% show no publication-free gap at all, while 27.3% show a gap of five years or more between active years. The five affiliations in Table 4 illustrate this variation: Tongji University was active in 12 of the 13 years since its first-indexed publication, with a longest gap of one year, whereas Southeast University was active in 10 of 14 years, with a longest gap of four years.

4.4 An integrative analytical framework: Participation–Persistence–Influence as dimensions of workforce configuration

The findings support a multidimensional interpretation of scientific workforce structure in which participation, longitudinal recurrence and scholarly influence are partially coupled but non-equivalent dimensions, valuable not for generating another researcher hierarchy but for revealing properties that conventional productivity rankings obscure. Three empirical patterns underpin this interpretation: participation diffusion, the broad distribution of entry into the research domain, with most author records contributing only once or twice; recurrence concentration, the substantially narrower distribution that emerges when attention shifts from entry to repeated publication across time — sharply so at institutional level, where 72.2% of affiliations publish in only one year despite distinct-document institutional output being comparatively dispersed (Gini = 0.39; HHI = 14/10,000; Section 3.3) — and productivity–influence decoupling, the incomplete correspondence between publication frequency and scholarly recognition.

These patterns build on, but differ from, established distinctions in the science-of-science literature: research on transient and continuing authors (Gordon, 2007), longitudinal heterogeneity in publication careers (Petersen et al., 2012; Milojević et al., 2018), and cumulative advantage in scientific recognition (Merton, 1968; Allison et al., 1982). The P–P–I framework connects these related but analytically separate observations at domain level by examining how broadly researchers participate, how much of that participation recurs, and whether scholarly influence follows the same distribution.

The Lotka goodness-of-fit result reported in Section 3.1 (β = 2.54 vs. the classical 2.0, Kolmogorov–Smirnov test p = .035) provides a first, genuine piece of quantitative distinctiveness evidence: this domain's productivity distribution is measurably more concentrated than the textbook Lotka approximation, not merely consistent with it. This claim is deliberately scoped to that comparison. The present dataset does not include a second, contemporary interdisciplinary domain analysed under the same method, so the broader question of whether this pattern of concentration is unusual relative to other current interdisciplinary materials fields is not addressed here and is not claimed. The contribution of the P–P–I framework, accordingly, is presented as a domain-specific empirical finding and an analytical distinction transferable to other fields, rather than as evidence that bentonite–cementitious research is distinctively unusual among contemporary research domains. The discrete maximum-likelihood re-estimate reported in Section 3.1 (α = 2.72–2.74) supports the same conclusion under a more efficient estimator, though a formal distributional comparison also found that a lognormal distribution fits the underlying productivity data better than a pure power law (Section 3.1), a nuance the Lotka framing conventionally used in bibliometrics does not capture.

The framework does not propose a developmental sequence from participation to persistence to influence; it provides a field-level structure for identifying their convergence and divergence. A domain may expand rapidly in contributor numbers while recurrent participation remains concentrated, recurrent contributors may not acquire equivalent influence, and influential contributors need not be the most prolific, differences that publication volume or citation rankings alone cannot establish.

4.5 Implications for interdisciplinary research fields

This distinction matters most for interdisciplinary fields. Bentonite–cementitious research draws on clay science, cementitious materials, geotechnical engineering and alternative binders, and such domains can attract contributors from neighbouring specialties without requiring them to remain continuously active in the focal literature.

A large contributor population should not, therefore, be read as evidence of equally extensive durable research capacity: a considerable share of the literature may come from episodic involvement, while continued activity rests on a much smaller recurrent population.

This has implications for research evaluation: rankings based solely on publication totals can overstate contributors whose output depends heavily on large-team participation, while rankings based solely on cumulative citations can favour longer-established researchers. Evaluating a field through multiple workforce dimensions gives a more defensible account of its organisation.

These implications remain bounded: bibliographic recurrence cannot establish whether recurrent contributors produce better technologies, whether cited authors are more influential in industrial adoption, or whether particular institutions merit greater investment, since such evidence lies outside bibliographic metadata. Growth in publications and contributor numbers should not, therefore, be treated as sufficient evidence of a broadly distributed, recurrent, influential research workforce.

4.6 Limitations

Several limitations affect interpretation of the workforce results. First, author-level analysis depends on bibliographic identity. Common surname–initial combinations may represent more than one researcher, while spelling and indexing variants can separate publications belonging to the same individual. Author labels should therefore be interpreted as bibliographic records unless independently disambiguated.

This concern is not merely illustrative. Parsing the per-document author-affiliation field (C1) in the underlying Bibliometrix export and clustering each candidate leading author record's documents by affiliation showed that of thirteen author records initially identified as the most productive or most cited in this corpus, three (Cuevas J, Chen Y, Plecas I) are consistent with a single individual, their documents cluster overwhelmingly on one affiliation and a stable full-name form, while the remaining ten are demonstrably composite (Strotmann & Zhao, 2012). For the record labelled “Li J,” the single most productive bibliographic label in the corpus (27 full-count publications), the largest group of documents attributable to one affiliation and full-name form is 8, meaning the true most productive individual sharing that label accounts for well under a third of the reported count; a comparable pattern (12–41% single-identity share) held across the other nine composite records (Online Resource 1). Because this contamination cannot currently be resolved to individually verified identities (disambiguation was checked for candidate leading-author records only, using affiliation clustering rather than ORCID-level verification), this manuscript does not report individual-level author rankings; Sections 3.2–3.4 and 4.2–4.3 accordingly report only population-level associations and institutional-level trajectories. Neither depends on resolving individual author identity. Online Resource 1 documents the full disambiguation check, including the corrected candidate top-author list, for readers who wish to evaluate it directly; a fully ORCID-verified re-analysis is left for future work. A sensitivity check addressing this concern directly is reported in Section 3.1: recalculating the Lotka exponent from the audited individual-level counts increases, rather than reduces, the estimated concentration among the thirteen audited leading records. Because only those thirteen records were audited, this check does not establish that the reported exponent is unaffected by identity errors elsewhere in the 4,752-record population; it establishes only that the disambiguation limitation does not explain away the concentration observed among the records actually checked.

Second, longitudinal recurrence does not represent a complete scientific career. Researchers may remain active outside the search boundary, experience intervals without publications, change research topics, or continue contributing in ways not captured by indexed outputs. The analysis consequently identifies recurrence within the retrieved literature rather than continuous employment or career persistence. Sections 3.3 and 4.3 report a more granular affiliation-level persistence measure (active years, active 5-year periods, and the longest publication-free gap; Online Resource 2) that does distinguish consecutive from intermittent recurrence; even so, this measure describes bibliographic presence, not the frequency or intensity of underlying research activity, and it is not available at author level because individual identities could not be reliably disambiguated (First limitation, above).

Third, citation influence is conditioned by publication age, disciplinary citation practices and database coverage. Local citations are additionally dependent on the boundary of the 1,492-document corpus. The h-, g- and m-indices provide complementary views of impact but cannot be interpreted as direct measures of research quality.

Fourth, the use of WoSCC and English-language publications limits corpus completeness. Relevant studies indexed elsewhere or published in other languages may not be represented, and restricting the corpus to the WoS Article and Review document types excludes conference proceedings papers, which may under-represent contributors whose participation in this domain is weighted toward conference venues rather than journal articles and reviews. The final observation year, 2026, is also incomplete because records were retrieved on 1 August 2026. The search string's precision was also not independently validated against a manually coded relevance sample; some retrieved records may fall outside the intended scope of bentonite–cementitious research, and relevant records using terminology outside the string may have been missed.

Fifth, two methodological choices in the productivity and concentration analyses warrant a qualified reading even after the additional checks reported in Sections 3.1 and 3.3. The Lotka exponent was initially estimated by OLS regression on log-transformed frequency data (Pao, 1985); Section 3.1 reports a discrete maximum-likelihood re-estimate (Clauset et al., 2009) that confirms the same conclusion under a more efficient estimator, but also shows, via a formal distributional comparison, that a lognormal model fits the productivity data better than a pure power law. The Lotka framing retained throughout this manuscript is therefore a conventional rather than a uniquely correct description of the productivity distribution. Section 3.3 reports the Gini coefficient, Herfindahl–Hirschman Index and top-k shares for institutional output using a distinct-document counting rule across all 1,674 affiliations (Online Resource 2). This avoids inflation from multiple co-authors sharing the same affiliation on one document and shows institutional output to be considerably less concentrated than author-level productivity. These additions narrow the class of claims that the productivity and concentration statistics can support rather than establishing a single uniquely correct distributional model.

Sixth, the Participation–Persistence–Influence framework is descriptive and analytical rather than causal. The study identifies relationships among bibliometric dimensions but cannot establish why researchers enter or leave the field, why some contributors become highly influential or whether longitudinal recurrence itself generates later citation impact.

5. Conclusion

This study examined the scientific workforce underlying bentonite–cementitious research by distinguishing three dimensions that are often conflated in bibliometric assessment: participation, longitudinal recurrence and scholarly influence. The findings show that these dimensions overlap but are not equivalent. Participation is broad, yet repeated contribution is concentrated among a substantially smaller proportion of the recorded author population. Full and fractionalised publication counts produce

different productivity profiles, while citation-based measures identify a partly different structure of scholarly influence. Longitudinal trajectories further show that recurrent contribution can take both long-duration and more recently concentrated forms.

The principal contribution of the study lies in demonstrating that the numerical expansion of a research field should not be equated with the breadth of its recurrent scientific workforce or the distribution of scholarly influence. A large contributor population can coexist with a comparatively narrow recurrent core, while prolific publication does not necessarily correspond to greater domain-specific influence. Scientific workforce structure is therefore better understood as multidimensional than as a hierarchy derived from publication counts or citation rankings alone. The distinction is level-dependent as well as dimension-dependent: distinct-document institutional output is comparatively dispersed, while institutional recurrence is concentrated among a much smaller subset of affiliations, a pattern that a single concentration statistic at one level of analysis would not reveal.

The proposed Participation–Persistence–Influence framework formalises this distinction at the analytical level. Participation captures the breadth and distribution of contribution; persistence, operationalised here as observable longitudinal recurrence, captures the continuity of bibliographic participation; and influence captures the scholarly visibility associated with that contribution. The framework does not imply a causal progression from participation to persistence and then to influence. Instead, it provides a basis for examining where these dimensions converge and where they diverge within a research domain.

For interdisciplinary fields, this distinction is consequential. Expansion may be supported by substantial entry from adjacent disciplines without producing an equally broad recurrent contributor base. Conversely, scholarly influence may become concentrated among researchers whose publication frequency does not place them among the most prolific contributors. Evaluating field development solely through publication growth, author counts or productivity rankings can therefore obscure important differences in the organisation of its knowledge-producing workforce.

These conclusions remain bounded by the characteristics of bibliographic data. Author records may contain unresolved homonyms or name variants, longitudinal recurrence does not represent complete scientific careers, and citation indicators remain sensitive to publication age, database coverage and the boundaries of the retrieved corpus. The findings should consequently be interpreted as evidence of bibliographically observable workforce structure, rather than as measures of employment continuity, researcher quality or causal processes of career development.

Within these boundaries, the study establishes a distinction with broader relevance to research evaluation: field size, recurrent scientific participation and scholarly influence are different properties of a research workforce and should be evaluated accordingly. Recognising this distinction provides a more defensible basis for assessing whether expansion in an interdisciplinary research domain is accompanied by recurrent participation and a broader distribution of scholarly influence.

Declarations

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval

Not applicable. This study analysed publicly available bibliographic metadata from the Web of Science Core Collection; it did not involve human or animal subjects.

Consent to Participate / Consent for Publication
Not applicable.

Funding

No funding source or grant; the research received no specific external funding.

Author Contribution

Ayokunle Odunayo Alade: Conceptualization, methodology, formal analysis, data curation, writing original draft. Shoaib Hassan: Writing - review & editing, Writing - original draft. Mineesha Sivakumar: Writing - review & editing, Writing - original draft. Shangeetha Ganesan: Writing - review & editing, Supervision. Mazidatulakmam Miskam: Writing - review & editing, Conceptualization, Supervision.

Acknowledgement

The authors acknowledge the staff and students at the School of Chemical Sciences, USM.

Data Availability

The bibliographic data were extracted from the Web of Science Core Collection under the authors' institutional licence and cannot be redistributed in raw form. Online Resource 1 contains the author-disambiguation audit and Lotka-exponent robustness checks. Online Resource 2 contains the institutional concentration and persistence dataset underlying Section 3.3. Further processed data tables are available from the corresponding author upon reasonable request.

References

1. Allison, P.D., Long, J.S., Krauze, T.K.: Cumulative advantage and inequality in science. Am. Sociol. Rev. **47**(5), 615–625 (1982). https://doi.org/10.2307/2095162
2. Aria, M., Cuccurullo, C.: bibliometrix: An R-tool for comprehensive science mapping analysis. J. Informetrics. **11**(4), 959–975 (2017). https://doi.org/10.1016/j.joi.2017.08.007
3. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. **51**(4), 661–703 (2009). https://doi.org/10.1137/070710111
4. 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). https://doi.org/10.1016/j.jbusres.2021.04.070
5. Egghe, L.: Theory and practise of the g-index. Scientometrics. **69**(1), 131–152 (2006). https://doi.org/10.1007/s11192-006-0144-7
6. Global Cement and Concrete Association (GCCA): Cement and Concrete Industry Net Zero Action and Progress Report 2025/26. GCCA (2025)
7. Gordon, A.: Transient and continuant authors in a research field: The case of terrorism. Scientometrics. **72**(2), 213–224 (2007). https://doi.org/10.1007/s11192-007-1714-z
8. Hasan, K., Jaya, R.P., Yahaya, F.M.: Application of bentonite in cement-based composites: A review of current status, challenges and future prospects. J. Building Eng. **98**, 111171 (2024). https://doi.org/10.1016/j.jobe.2024.111171
9. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. U.S.A. **102**(46), 16569–16572 (2005). https://doi.org/10.1073/pnas.0507655102
10. International Energy Agency (IEA): Breakthrough Agenda Report 2025: Cement and Concrete. IEA (2025)
11. Lotka, A.J.: The frequency distribution of scientific productivity. J. Wash. Acad. Sci. **16**(12), 317–323 (1926)
12. Merton, R.K.: The Matthew effect in science. Science. **159**(3810), 56–63 (1968). https://doi.org/10.1126/science.159.3810.56
13. Milojević, S., Radicchi, F., Walsh, J.P.: Changing demographics of scientific careers: The rise of the temporary workforce. Proc. Natl. Acad. Sci. U.S.A. **115**(50), 12616–12623 (2018). https://doi.org/10.1073/pnas.1800478115
14. Pao, M.L.: Lotka's law: A testing procedure. Inf. Process. Manag. **21**(4), 305–320 (1985). https://doi.org/10.1016/0306-4573(85)90055-X
15. Petersen, A.M., Riccaboni, M., Stanley, H.E., Pammolli, F.: Persistence and uncertainty in the academic career. Proc. Natl. Acad. Sci. U.S.A. **109**(14), 5213–5218 (2012). https://doi.org/10.1073/pnas.1121429109
16. Strotmann, A., Zhao, D.: Author name disambiguation: What difference does it make in author-based citation analysis? J. Am. Soc. Inform. Sci. Technol. **63**(9), 1820–1833 (2012). https://doi.org/10.1002/asi.22695

17. Waltman, L., van Eck, N.J.: Field-normalized citation impact indicators and the choice of an appropriate counting method. J. Informetrics. **9**(4), 872–894 (2015). https://doi.org/10.1016/j.joi.2015.08.001
18. Wu, L., Wang, D., Evans, J.A.: Large teams develop and small teams disrupt science and technology. Nature. **566**(7744), 378–382 (2019). https://doi.org/10.1038/s41586-019-0941-9

Figures

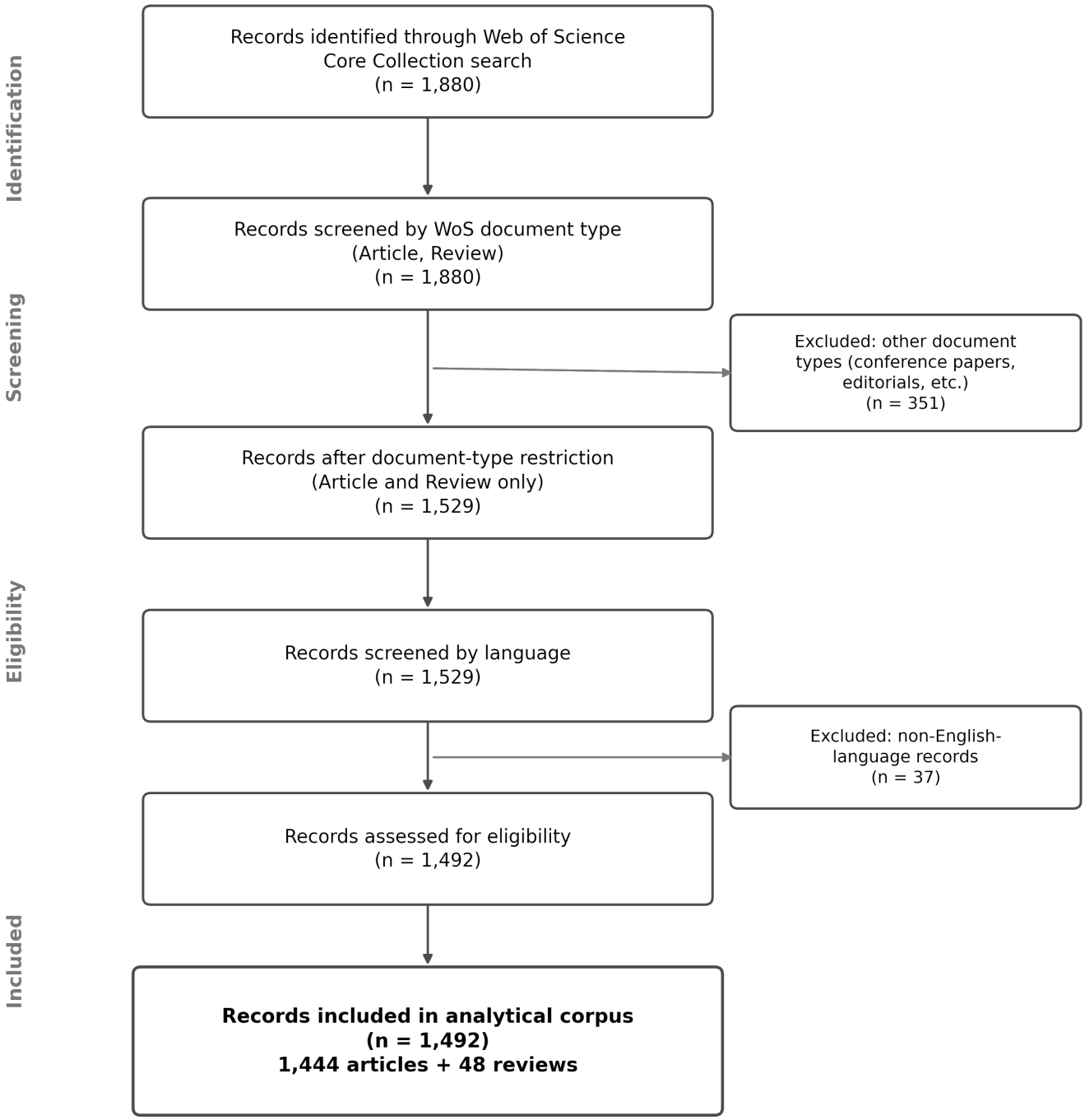


Figure 1

PRISMA-style flow diagram of corpus construction

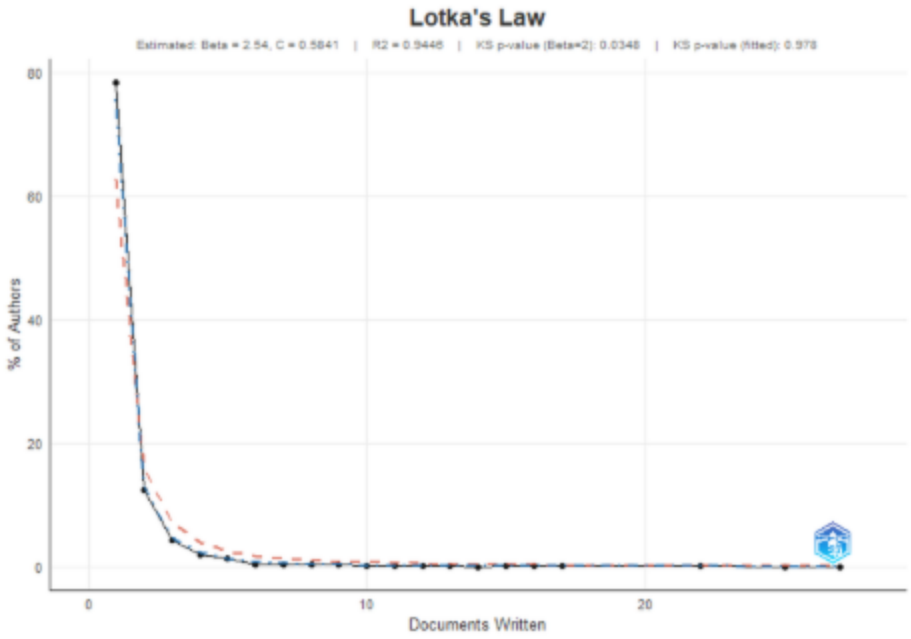


Figure 2

Distribution of author productivity in bentonite–cementitious research

Supplementary Files

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

- OnlineResource1QQ.xlsx
- OnlineResource2QQ.xlsx