Errors in documenttype classification: a focus on engineering publications and their publishers

Scientometrics (2025) 130:6613–6641
https://doi.org/10.1007/s11192-025-05444-6

Errors in document-type classification: a focus on engineering publications and their publishers

Domenico A. Maisano[1] · Lucrezia Ferrara[1] · Fiorenzo Franceschini[1]

Received: 29 May 2025 / Accepted: 3 October 2025 / Published online: 16 October 2025

Abstract

This study investigates document-type (DT) classification errors—such as misclassifications of *research articles*, *reviews*, *conference proceedings*, *editorials*, etc.—in the bibliometric databases Scopus and Web of Science, focusing on the field of *engineering*. Such misclassifications can adversely affect academic research and research quality assessments, with potential repercussions on researchers' careers and the allocation of funding in academic institutions. By analysing a *corpus* of approximately 10,000 publications, a semi-automated approach is employed to identify misclassified documents and attribute errors to the responsible database. Additionally, the role of publishers (e.g., "Elsevier", "Springer Nature", "Taylor & Francis", etc.) is investigated, based on the hypothesis that certain publishers may contribute more significantly than others to DT-classification errors, due to specific editorial practices or inconsistencies in metadata. The results reveal that classification errors are non-negligible (i.e., they occur at rates of a few percentage points) and that the extent of publishers' contributions varies significantly in both Scopus and WoS. However, the most problematic publishers for each database appear to be uncorrelated. Integrating various statistical tests, this study provides insights that may be valuable to researchers, research evaluators, database operators, and publishers, raising awareness of the issue and offering preliminary indications for identifying possible remedies to mitigate such errors and enhance DT-classification accuracy. Future research will further investigate the reasons behind the concentration of errors among certain publishers.

Keywords Document-type classification · Classification error · Bibliometric databases · Engineering publications · Publisher influence

Introduction

Document types (DTs) are labels used to classify scientific documents based on their nature (e.g., *research articles*, *reviews*, *conference proceedings*, *letters*, *editorials*, etc.). These classifications may help target bibliographic searches, enabling users to retrieve relevant

✉ Fiorenzo Franceschini
fiorenzo.franceschini@polito.it

[1] Department of Management and Production Engineering (DIGEP), Politecnico di Torino, Turin, Italy

documents more effectively (Donner, 2017; Harzing, 2013; Yeung, 2021). DTs may be assigned by various entities, including the authors themselves, editorial committees, publishers, and multiple bibliometric databases (Clarivate 2025; Elsevier 2025). However, it is not rare for the same document to be classified differently by these entities.

This study focuses on the classifications performed by the two leading generalist bibliometric databases, Scopus and Web of Science (WoS), which are widely used in the scientific community, not only for document retrieval, but also for bibliometric analyses. These databases are frequently employed in research output evaluations for individual researchers, groups or institutions, in contexts such as academic promotions, hiring processes and national research assessment exercises (García-Pérez, 2010). In such evaluation scenarios, DT-classification errors can have significant practical consequences, including the erroneous exclusion of documents from assessments, the downgrading of documents (e.g., a *research article* labelled as a *review* or *conference proceeding*), or, conversely, the undue promotion of other documents (e.g., a *short communication* to a *research article*) (Mokhnacheva, 2023; Sigogneau, 2000; Yeung, 2019).

The scientific literature includes some studies on DT-classification errors in bibliometric databases, although these are generally limited in terms of the number of analysed articles (Donner, 2023; Elango, 2025; Franceschini et al., 2013, 2015, 2016a; García-Pérez, 2010; Moed, 2005; Olensky et al., 2016; Valderrama-Zurián et al., 2015). A common finding among these studies is that, for both generalist databases, DT-classification errors occur at a rate of a few percentage points—which is low but not negligible (Elango, 2025; Haupka et al., 2024).

Recently, Maisano et al. (2025a) introduced a new semi-automated methodology for investigating DT-classification errors, based on the comparison of discrepant classifications between Scopus and WoS for the same set of documents. This methodology is structured in two phases: the first (automated) phase identifies documents potentially affected by classification errors; the second (manual) phase examines these documents to attribute the errors to the responsible database. By limiting the inherently labour-intensive and time-consuming manual analysis to a critical subset of documents, this approach significantly enables the examination of tens/hundreds of thousands of documents instead of the hundreds/thousands typically addressed in previous studies. This methodology was recently applied in a study analysing ~ 30,000 documents produced by faculty members at the two leading universities in Turin (Politecnico di Torino and University of Turin) over the five-year period from 2019 to 2023.

The present research builds on that investigation, focusing on a subset of documents produced by faculty at the Politecnico di Torino (hereafter abbreviated as PoliTO), a technical university in Italy specializing in *engineering* and *architecture*. The analysis of this *corpus* allowed us not only to quantify DT-classification errors in Scopus (error rate ~ 2.1%) and WoS (error rate ~ 1.3%), as detailed in sect. "Methodology" (Maisano et al., 2025a), but also to explore broader patterns potentially underlying these errors. In particular, this study investigates whether certain publishers are systematically associated with a higher incidence of DT-classification errors. Building on the hypothesis that publishers' editorial policies or metadata practices may play a role in these discrepancies, we address the following research questions:

RQ1 *Is the publisher a systematic factor in DT-classification errors? Are there significant differences in the propensity to make these errors depending on the publisher?*

RQ2 *Are there similarities or differences in the distribution of errors among publishers across the two databases?*

By addressing these questions, the study seeks to provide a deeper understanding of the interplay between publishers and bibliometric databases in the generation of DT-classification errors. A structured methodological approach is adopted, consisting of several steps, summarised below:

- Document *corpus* identification: Selection of a large set of scientific documents in the field of *engineering*. An aggregated study is preferred over a focus on individual subfields (e.g., *mechanical*, *civil*, *manufacturing*, *electrical engineering*), given the relative homogeneity of publishing practices in this area (Franceschini & Maisano, 2014).
- Publisher identification: Linking each document to its respective publisher.
- Publisher rationalisation: Aggregation of the many different publishers into groups of publishers, to reduce fragmentation and simplify analysis; e.g., consolidating "Springer", "Springer London" and "Palgrave Macmillan" into a single macro-publisher "Springer Nature", as will be shown in detail in sect. "Identification and rationalisation of publishers".
- Detection of classification errors: Identification of DT-classification errors and assignment of responsibility to the relevant database. Calculation of error statistics for both databases.
- Focus on publishers: Assignment of individual database errors to the relevant macro-publishers.
- Statistical analysis: Separate statistical analysis for each database, aimed at evaluating systematic differences between publishers in their propensity for DT-classification errors (cf. RQ1).
- Database comparison: Comparison of results from the two databases to identify potential similarities and differences in the observed patterns (cf. RQ2).

The remainder of this article is organised into four sections. Section "Methodology" provides a detailed explanation of the methodological approach, consistent with the steps outlined above. Section "Results" presents an in-depth description of the results, interpreting them in relation to the research questions. The conclusions summarize the original contributions of this study, highlighting practical implications, limitations, and potential directions for future research. Finally, the (online) appendix contains further insights and details on the results of the analysis, from the perspective of the individual publishers.

Methodology

This section outlines the methodological approach adopted in the analysis, divided into three subsections, each focusing on a key phase:

- Data collection and identification of DT-classification errors and relevant statistics, for both databases;
- Identification of publishers and their aggregation into macro-publishers, so that the analysis is not fragmented;

Table 1 *Concordance matrix* for the *corpus* of interest of 10,834 documents

DT classifications →		by Scopus									
↓		*Article*	*Conf. paper*	*Review*	*Editorial*	*Book chapter*	*Erratum*	*Note*	*Letter*	*Short Survey*	**Row total**
by WoS	*Article*	(8,378)	29	100	3	3	–	10	6	2	**8,531**
	Proceedings Paper	15	(1,463)	–	1	55	–	–	–	–	**1,534**
	Review	70	–	(413)	–	–	–	–	–	3	**486**
	Editorial Material	20	4	17	(162)	–	1	14	–	3	**221**
	Correction	1	–	–	–	–	(48)	–	–	–	**49**
	Letter	2	–	–	–	–	–	–	(7)	–	**9**
	Retraction	–	–	–	–	–	2	–	–	–	**2**
	Meeting Abstract	–	1	–	–	–	–	–	–	–	**1**
	Book Review	1	–	–	–	–	–	–	–	–	**1**
	Column total	**8,487**	**1,497**	**530**	**166**	**58**	**51**	**24**	**13**	**8**	**10,834**

While the elements in the diagonal (in round brackets) denote DT classifications that are concordant between competing databases, those off-diagonal denote potential DT-classification errors. Documents corresponding to the latter elements were manually analysed

- Analysis of error statistics relating to specific publishers, considering each database separately.

Data collection and error statistics

Building on a previous study (Maisano et al., 2025a), a subset of over ten thousand scientific documents produced by the (structured) academic staff of PoliTO[1] in the five-year period 2019–2023 was analysed. The selection of this *corpus* of documents focuses the investigation to the field of *engineering*, a broad area characterised by relatively homogeneous publishing practices across specialties. This homogeneity makes the study relevant to a significant portion of the scientific community, while avoiding the specificity of individual sub-disciplines (Franceschini & Maisano, 2014).

Among the publications issued during the five-year period from 2019 to 2023, precisely 10,834 were selected, indexed by both bibliometric databases, Scopus and WoS, and provided with DOI codes. This selection is necessary to apply the semi-automated technique proposed by Maisano et al., (2025a, 2025b), which—based on the comparison of potentially discrepant DT classifications between the two databases—enables the automatic identification of a subset of "discordant" documents, which are then subjected to further manual analysis. Table 1 shows the so-called *concordance matrix* resulting from this initial phase of analysis, with the main DTs assigned by Scopus to the individual documents in the *corpus*, listed in the columns, and those assigned by WoS to the same documents, listed in the rows. The "concordant" documents, which are 10,471 out of 10,834 in all, are those located on the diagonal, in brackets "(…)", and are defined as such because they have concordant DT classifications between the two databases (e.g., a *conference paper* in Scopus and a *conference proceeding* in WoS). While previous studies have documented cases of DT-classification errors, including concurrent misclassifications by Scopus and WoS under specific conditions (Zhu et al., 2024), the overall likelihood of such errors among these documents is very low and therefore they are subjected to manual analysis only on a sample basis, as explained in more detail below (Maisano et al., 2025a). Conversely, documents located off the diagonal are classified as "discordant" as they display discordant DT

[1] PoliTO is Italy's second-largest technical university, with approximately a thousand faculty members (including *full*, *associate*, and *assistant* professors), serving a student body of ~ 38,000 people, primarily in *engineering*-related fields.

classifications between the two databases (e.g., *research article* versus *review* or *letter*). These (363) discordant documents are highly likely to include a DT-classification error attributable to one of the two databases. For this reason, they are subjected to 100% manual analysis.

In this specific case, a sample of 2% of the 10,471 "concordant" documents, corresponding to 210 documents, were manually analysed, without detecting any DT-classification errors and thus confirming the hypothesis that the probability of errors among these documents is minimal (Maisano et al., 2025a, 2025b). On the other hand, the manual analysis of all 363 "discordant" documents revealed a significant number of DT-classification errors attributable to one or sometimes both databases, with responsibility in rarer cases shared between the two.

The so-called *error tables* in Tables 2 and 3 summarize the analysis results from the perspectives of Scopus and WoS respectively. The rows represent the "true" DTs assigned to the documents, as determined through manual analysis, while the columns indicate actual DTs assigned by the database of interest. Each error table visualises the distribution of errors across the various DT categories. The diagonal elements in brackets "(…)" represent correct assignments, DT by DT, while the off-diagonal elements indicate errors of various types. For example, analysing the rows reveals *false exclusions* from the DT indicated in the column. Conversely, analysing the columns highlights *false inclusions*, where documents are mistakenly classified into a DT different from their actual one. In the subsequent part of the section, specific statistics are introduced to quantify these types of errors.

As documented in previous studies, one of the most common errors involves the reciprocal misclassification of *reviews* and *research articles*, observed in both databases. For example, considering Table 2, 28 out of 530 documents are (mis)classified by Scopus as *reviews* instead of as *articles*. Conversely, in Table 3, 63 out of 8,531 documents are erroneously classified by WoS as *articles* instead of *reviews*. Another typical error involves the undue downgrading of a research *article* to *conference paper* or vice versa. Considering Table 2, 23 out of 1,497 *conference papers* are erroneously classified by Scopus as *articles.* Additionally, Table 2 reveals a relatively large amount of documents (mis)classified as *book chapters* by Scopus instead of *conference papers* (i.e., 51 of the 58 documents classified as *book chapters* are incorrect).

From any error table, the most general error statistic to represent the overall accuracy of the database of interest is:

$$\varepsilon = \frac{\sum_{\forall(i,j)|i\neq j} x_{(i,j)}}{\sum_{\forall(i,j)} x_{(i,j)}}, \tag{1}$$

where $x_{(i,j)}$ is the number of documents reported in the i-th row and j-th column of the error table, for the database of interest. The ε values resulting from the analysis are shown in the lower-right corner of the error tables. It can be observed that for both databases, ε remains within the range of a few percentage points, with slightly higher values for Scopus (2.1%) compared to WoS (1.3%). These error statistics, although restricted to the field of *engineering*, align with those reported in previous studies conducted on more heterogeneous document datasets (Haupka et al., 2024; Maisano et al., 2025a; Mokhnacheva, 2023).

Next, two other more detailed database-error statistics can be constructed from the perspective of a specific DT (Maisano et al., 2025a). The first one (α_i) expresses the probability that a document belonging to a specific (i-th) DT is wrongly classified into a different DT (i.e., *false exclusion from the DT of interest*):

Table 2 *Error table* for Scopus, with error statistics (see Eqs. 1, 2, and 3) highlighted in bold

		DT classification by Scopus										
		Article	Conf. Paper	Review	Editorial	Book chapter	Erratum	Note	Letter	Short Survey	Row total	α_i
"True" DT classifications	Article	(8,400)	23	28	–	3	–	2	–	1	8,457	**0.7%**
	Conf. Paper	15	(1,473)	–	1	51	–	–	–	–	1,540	**4.4%**
	Review	50	–	(483)	–	–	–	–	–	1	534	**9.6%**
	Editorial	14	1	14	(164)	–	–	3	–	3	199	**17.6%**
	Erratum	1	–	–	–	–	(49)	–	–	–	50	**2.0%**
	Note	5	–	2	1	–	–	(19)	–	1	28	–
	Letter	1	–	–	–	–	–	–	(13)	–	14	–
	Book Chapter	–	–	–	–	(4)	–	–	–	–	4	–
	Other	1	–	3	–	–	–	–	–	–	4	–
	Short Survey	–	–	–	–	–	–	–	–	(2)	2	–
	Retracted	–	–	–	–	–	2	–	–	–	2	–
	Column total	8,487	1,497	530	166	58	51	24	13	8	10,834	
	β_j	**1.0%**	**1.6%**	**8.9%**	**1.2%**	**93.1%**	**3.9%**	–	–	–		$\varepsilon \cong$ **2.1%**

The diagonal elements indicating correctly classified documents are marked with round brackets. The a_{DT} and b_{DT} statistics (in the right-hand column and bottom row respectively) are only calculated for subgroups with at least 30 documents, to ensure statistical robustness

Table 3 *Error table* for WoS, with error statistics (see Eqs. 1, 2, and 3) highlighted in bold

	DT classification by WoS											
		Article	Proc. Paper	Review	Editorial Material	Correction	Letter	Retraction	Book Review	Meeting Abstract	Row total	α_i
"True" DT classifica-tions	Article	(8,441)	–	17	1	–	1	–	–	–	8,460	**0.2%**
	Proc. Paper	6	(1530)	–	3	–	–	–	–	1	1,540	**0.7%**
	Review	63	–	(467)	2	–	–	–	–	–	532	**12.2%**
	Editorial Material	3	–	–	(198)	–	–	–	–	–	201	**1.5%**
	Correction	–	–	–	1	(49)	–	–	–	–	50	**2.0%**
	Other	8	–	2	15	–	–	–	–	–	25	–
	Letter	6	–	–	–	–	(8)	–	–	–	14	–
	Book Chapter	–	4	–	–	–	–	–	–	–	4	–
	News Item	3	–	–	–	–	–	–	–	–	3	–
	Retraction	–	–	–	–	–	–	(2)	–	–	2	–
	Book Review	1	–	–	–	–	–	–	(1)	–	2	–
	Biographical-Item	–	–	–	1	–	–	–	–	–	1	–
	Column total	8,531	1,534	486	221	49	9	2	1	1	10,834	
	β_j	**1.1%**	**0.3%**	**3.9%**	**10.4%**	**0.0%**	–	–	–	–		$\varepsilon \cong$ **1.3%**

The diagonal elements indicating correctly classified documents are marked with round brackets. The a_{DT} and b_{DT} statistics (in the right-hand column and bottom row respectively) are only calculated for subgroups with at least 30 documents, to ensure statistical robustness

$$\alpha_i = \frac{\sum_{\forall(i,j)|j\neq i} x_{(i,j)}}{\sum_{\forall(i,j)} x_{(i,j)}}, \tag{2}$$

α_i is calculated on a row-by-row basis by means of the ratio of misclassifications to the row total. This indicator can be interpreted—with reference to a specific row of the error table—as the ratio of the sum of the numerical values outside the parentheses to the row total.

The second database-error statistic (β_j) expresses the probability that a document is misclassified into a specific (j-th) DT of interest (i.e., *false inclusion into the DT of interest*):

$$\beta_j = \frac{\sum_{\forall(i,j)|i\neq j} x_{(i,j)}}{\sum_{\forall(i,j)} x_{(i,j)}} \tag{3}$$

β_j is calculated on a column-by-column basis by means of the ratio of misclassifications to the column total. This indicator can be interpreted—with reference to a specific column—as the ratio of the sum of the numerical values outside the parentheses to the column total. To avoid providing non-robust statistics, α_i and β_j are only calculated for subgroups of at least 30 documents.

Identification and rationalisation of publishers

This phase of the analysis considers the more granular level of *publishers*, as organizations responsible for the editorial aspects of scientific documents (Csomós & Farkas, 2023). These editorial aspects are generally distinct from the scientific content, which – instead – is typically managed by scientific committees composed of experts and supported by peer reviewers (Franceschini et al., 2014). For each document of interest, publisher information was extracted using the specific "Publisher" field in Scopus and WoS (Baas et al., 2020).

The analysis revealed a highly fragmented scenario, characterised by a huge number of (apparently) distinct publishers: as many as 333 total publishers emerged from the union of those associated with the documents of interest! This fragmentation is attributable not only to the existence of many genuinely distinct publishers but also to the complex structure of the major publishing groups, which include numerous sub-publishers, often inconsistently reported by bibliometric databases. For example, Elsevier's numerous divisions include "Elsevier GmbH", "Elsevier USA", "Elsevier Masson s.r.l.", "Elsevier B.V.", "Academic Press", "Academic Press Inc." and many others, as explained in more detail below. Additionally, significant heterogeneity was observed in the naming of the same (sub-)publisher across different documents (e.g., "Blackwell Publishing", "Blackwell Publishing Inc.", "Blackwell Publishing Ltd" or "John Wiley and Sons Inc", "John Wiley and Sons Inc.", "John Wiley and Sons Ltd", "John Wiley and Sons Ltd.", etc.).

To address this fragmentation and heterogeneity in the naming of publishers, a preliminary operation of data cleaning, rationalisation and aggregation was carried out. This process involved grouping sub-publishers belonging to the same editorial groups into "macro-publishers". The underlying rationale is that macro-publishers typically impose standardized styles, editorial formats, work practices and, likely, conventions on the documents managed by their internal (sub-publisher) divisions (Cambridge University Press, 2023). Moreover, some level of homogeneity is expected in the metadata provided to bibliometric databases, to facilitate document indexing (Crotty, 2024). As a result, treating all documents associated with the same macro-publisher as a homogeneous group enables the creation of larger samples, enhancing the robustness of statistical analyses.

The aggregation into macro-publishers was performed manually, using online information about the publishers of the documents and their affiliation with broader editorial groups (Larivière et al., 2015; van Bellen et al., 2024). Table A.55 (in online Appendix A.3) exemplifies the aggregation process of some publishers into their corresponding macro-publishers, detailing the sources and criteria used for this aggregation. This rationalisation resulted in the creation of 218 macro-publishers. To prevent excessive fragmentation in the analysis, only macro-publishers with at least 30 documents were considered, as summarized in Table 4. A total of 26 macro-publishers met this criterion, while the remaining ones were grouped under the category "*Others*", as their relatively low document count makes them statistically less significant. To simplify the discussion, from now on the term "publishers" will be used to refer to these 26 macro-publishers.

The Pareto chart in Fig. 1 summarises the results of the rationalisation process, which also indicates the number of documents indexed and the corresponding percentage of discordant ones, from the perspective of each of the 26 macro-publishers. This information provides a preview of the potentially most problematic publishers from the point of view of both databases, as shown in more detail in sect. "Results". The cumulative chart in Fig. 2 shows that the top 26 macro-publishers listed in Table 4 account for approximately 94% $\left(\text{i.e., } \frac{\textit{Documents of the major 26 macro-publishers}}{\textit{Total corpus of documents}} = \frac{10{,}169}{10{,}834} \approx 94\%\right)$ of the total documents included in the analysis (left scale) and 88% $\left(\text{i.e., } \frac{\textit{Discordant documents of the major 26 macro-publishers}}{\textit{Discordant documents in the total corpus}} = \frac{320}{363} \approx 88\%\right)$ of the identified discordant documents (right scale). This result aligns with the simplifying choice of excluding macro-publishers with fewer documents from the analysis and grouping them into the "*Others*" category.

Publisher-specific analysis for single databases and subsequent comparison

The scheme in Fig. 3 illustrates how the "overall" error table of a database and related error statistics can be decomposed into publisher-specific error tables and related error statistics attributable to each p-th publisher. The overall error rate (ε, cf. Equation 1) of a database can be interpreted as a weighted sum of the error contributions (ε_p) related to the documents from each p-th publisher:

$$\varepsilon = \frac{E}{D} = \frac{\sum_{\forall p} e_p}{\sum_{\forall p} d_p} = \frac{\sum_{\forall p} \varepsilon_p \cdot d_p}{\sum_{\forall p} d_p}, \tag{4}$$

where.

E is the total number of documents with DT-classification errors;

D is the total number of documents of interest;

d_p is the total number of documents for the p-th publisher;

e_p is the number of observed documents with DT-classification errors for the p-th publisher;

$\varepsilon_p = \frac{e_p}{d_p}$ is the error rate for the p-th publisher.

Once the error statistics for each p-th database have been determined, targeted statistical tests (χ^2 and pairwise-difference tests in error rates) can be conducted to identify significant differences between documents from different publishers, in terms of their propensity

Table 4 Results of the rationalisation of publishers

Macro-publishers and their corresponding sub-publishers	No. of documents
(A) Elsevier	**2,735**
Academic Press	89
Academic Press Inc	61
Acta Materialia Inc	10
Associazione Italiana di Fisica Medica	3
Australian Society for Parasitology	1
Biophysical Society	3
Cell Press	1
Desalination Publications	2
Elsevier	1
Elsevier B.V	893
Elsevier Editora Ltda	5
Elsevier GmbH	9
Elsevier Inc	73
Elsevier Ireland Ltd	31
Elsevier Ltd	1,503
Elsevier Masson s.r.l	18
Elsevier Masson SAS	17
Elsevier S.A	1
Elsevier USA	10
Mosby Inc	3
Neoplasia Press, Inc	1
(B) MDPI	**1,812**
MDPI	779
MDPI AG	1,013
MDPI Multidisciplinary Digital Publishing Institute	20
(C) IEEE	**1,648**
Education Society of IEEE (Spanish Chapter)	1
IEEE Canada	2
IEEE Computer Society	157
Institute of Electrical and Electronics Engineers	4
Institute of Electrical and Electronics Engineers Inc	1,484
(D) Springer Nature	**1,290**
Adis	2
ALTEX Edition	2
BioMed Central Ltd	13
BioMed Central Ltd	10
Birkhauser	22
Joint Center on Global Change and Earth System Science of the University of Maryland and Beijing Normal University	1
Nature Publishing Group	42
Nature Research	99
Palgrave Macmillan	1
Springer	242
Springer Basel	1

Table 4 (continued)

Macro-publishers and their corresponding sub-publishers	No. of documents
Springer Berlin Heidelberg	4
Springer Boston	1
Springer International Publishing	25
Springer Japan	3
Springer London	7
Springer Medizin	1
Springer Nature	17
Springer Netherland	1
Springer Netherlands	62
Springer New York LLC	89
Springer Science and Business Media B.V	92
Springer Science and Business Media Deutschland GmbH	362
Springer Science and Business Media, LLC	9
Springer Science + Business Media B.V	4
Springer Verlag	154
SpringerOpen	3
Springer-Verlag France	4
Springer-Verlag Italia s.r.l	9
Springer-Verlag Wien	6
Wuhan Ligong Daxue	2
(E) John Wiley & Sons, Inc	**515**
Blackwell Publishing	2
Blackwell Publishing Inc	11
Blackwell Publishing Ltd	81
Equine Veterinary Journal Ltd	1
Hindawi Limited	54
John Wiley and Sons Inc	153
John Wiley and Sons Inc	18
John Wiley and Sons Ltd	94
John Wiley and Sons Ltd	3
Statistical Society of Canada	1
Wiley Blackwell	3
Wiley-Blackwell	9
Wiley-Liss Inc	6
Wiley–VCH Verlag	79
(F) Taylor & Francis Group	**321**
Bellwether Publishing, Ltd	25
CRC Press/Balkema	11
Dove Medical Press Ltd	1
Future Medicine Ltd	7
Routledge	51
Taylor and Francis Inc	49
Taylor and Francis Ltd	7
Taylor and Francis Ltd	170

Table 4 (continued)

Macro-publishers and their corresponding sub-publishers	No. of documents
(G) IOP Publishing Ltd	**308**
Institute of Physics	31
Institute of Physics Publishing	148
IOP Publishing Ltd	126
Web Portal IOP	3
(H) ACS Publications	**179**
American Chemical Society	179
(I) SAGE Publications	**176**
IOS Press	14
IOS Press BV	4
SAGE Publications Inc	24
SAGE Publications Ltd	134
(J) American Physical Society (APS)	**172**
American Physical Society	172
(K) Frontiers Media S.A	**141**
Frontiers Media S.A	141
(L) AIP Publishing LLC	**117**
Acoustical Society of America	8
American Association of Physics Teachers	2
American Institute of Physics Inc	105
AVS Science and Technology Society	2
(M) Royal Society of Chemistry	**104**
Royal Society of Chemistry	104
(N) Association for Computing Machinery, Inc (ACM) Publications	**83**
Association for Computing Machinery	55
Association for Computing Machinery, Inc	28
(O) Emerald Group Publishing Ltd	**76**
Emerald Group Holdings Ltd	36
Emerald Group Publishing Ltd	5
Emerald Publishing	9
ICE Publishing	26
(P) SPIE	**64**
SPIE	64
(Q) Optica Publishing Group	**56**
Optica Publishing Group (formerly OSA)	5
OSA—The Optical Society	34
The Optical Society	17
(R) American Society of Mechanical Engineers (ASME)	**51**
American Society of Mechanical Engineers (ASME)	51
(S) Cambridge University Press	**45**
Cambridge University Press	45
(T) World Scientific Publishing Co. Pte Ltd	**45**
World Scientific	22
World Scientific Publishing Co. Pte Ltd	23

Table 4 (continued)

Macro-publishers and their corresponding sub-publishers	No. of documents
(U) American Society of Civil Engineers (ASCE)	**44**
American Society of Civil Engineers (ASCE)	44
(V) Copernicus Publications	**42**
Copernicus GmbH	36
Copernicus Publications	6
(W) De Gruyter Brill	**40**
Birkhauser Boston	3
Birkhauser Verlag AG	7
De Gruyter	3
De Gruyter Open Ltd	17
Sciendo	9
Walter de Gruyter GmbH	1
(X) Oxford University Press	**39**
Oxford University Press	39
(Y) Science Press Ltd	**36**
Chinese Academy of Sciences	2
EDP Sciences	34
(Z) Public Library of Science	**30**
Public Library of Science	30
Others	
(192 macro-publishers with < 30 documents each)	**665**

Macro-publishers (bolded) are listed in descending order based on the overall number of documents included in the analysis. Macro-publishers with fewer than 30 total documents were grouped into the "*Other*" category

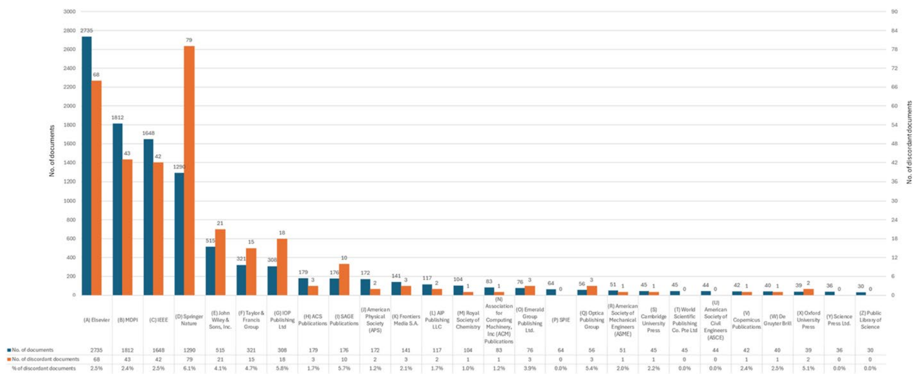


Fig. 1 Pareto chart of the 26 macro-publishers with higher number of documents. For each macro-publisher, the total number of documents (left-hand scale) and the total number of discordant documents (right-hand scale) between Scopus and WoS are shown at the top of the corresponding bar

for DT-classification errors. Subsequently, the statistical analysis can be extended to explore the similarities and differences in the results concerning the two databases.

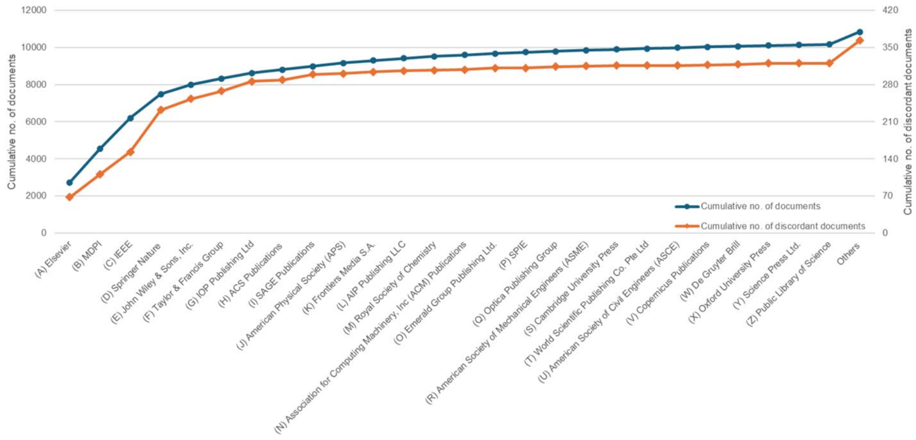


Fig. 2 Cumulative diagram of the total number of documents considered in the analysis (left-hand scale) and the total number of discordant documents (right-hand scale). The macro-publishers on the *x*-axis are listed in descending order based on the total number of documents included in the analysis (cf. Table 4)

Results

This section is organised into three subsections. The first one examines the results of the analysis for the Scopus database, the second one for the WoS database and the third one provides a structured comparison aimed at highlighting similarities and differences between the two databases. Analyses are supported by various statistical tests.

Publisher-specific analysis for Scopus

The results of the publisher-specific analysis for Scopus are summarized in Table 5. For each p-th publisher of the major 26 publishers and the "*Others*" category (listed in the table rows; cf. Table 4 in sect. "Identification and rationalisation of publishers"), d_p, e_p, the corresponding e_p (cf. Equation 4), and the 95% confidence interval (CI) are computed as detailed below (see Eqs. 6 and 8).

The error statistics for documents associated with a given p-th publisher can be modelled as a binomially distributed variable, with parameters d_p and ε_p, characterised by the following mean (m) and variance (s^2):

$$E_p \sim Bin\left[\mu = d_p \cdot \varepsilon_p, \sigma^2 = d_p \cdot \varepsilon_p \cdot \left(1 - \varepsilon_p\right)\right] \tag{5}$$

To estimate ε_p and provide a 95% confidence interval for the error probability, the *Wilson score interval*, which provides better coverage and accuracy when dealing with small sample sizes or proportions close to 0 or 1, can be calculated as (Devore, 2016):

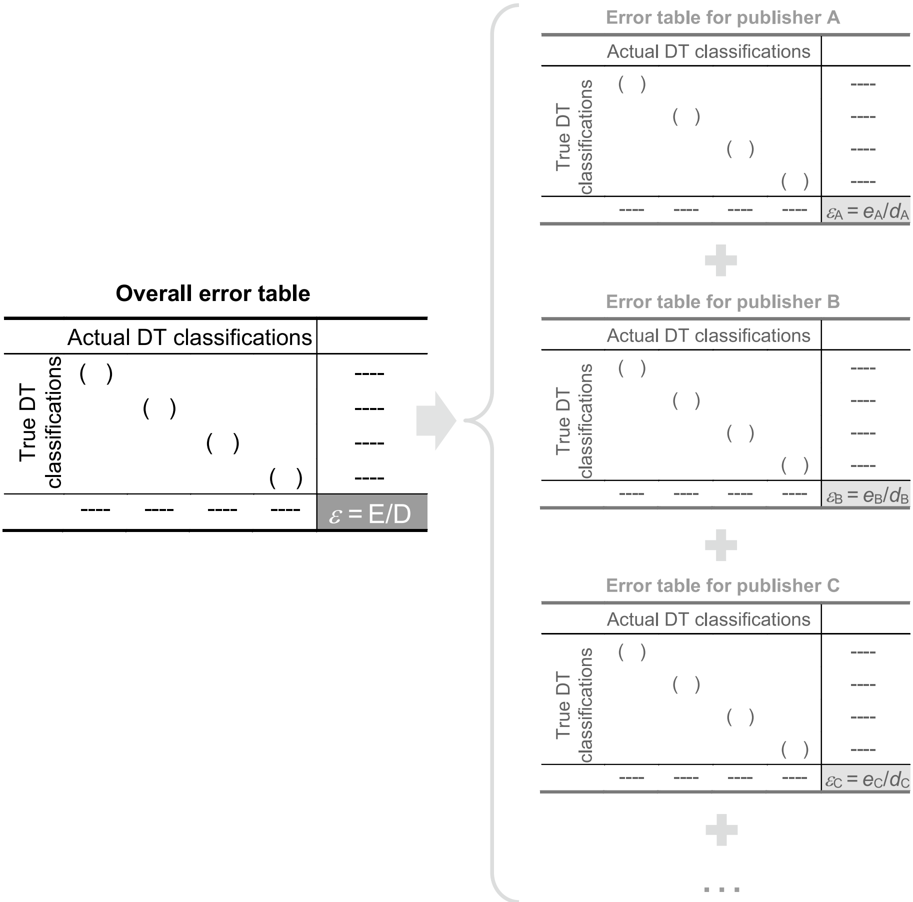


Fig. 3 Schematic representation of the decomposition of the overall error table of a database into relative publisher-specific error tables, according to a (weighted) additive logic (see Eq. 4). See online Appendix A.1 for Scopus error tables and online Appendix A.2 for WoS error tables related to the major 26 macro-publishers

$$\varepsilon_p = \frac{\hat{\varepsilon}_p + \frac{z^2_{1-\frac{\alpha}{2}}}{2 \cdot d_p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\varepsilon}_p \cdot (1-\hat{\varepsilon}_p)}{d_p} + \frac{z^2_{1-\frac{\alpha}{2}}}{4 \cdot d_p^2}}}{1 + \frac{z^2_{1-\frac{\alpha}{2}}}{d_p}} \tag{6}$$

where:

α is the type-I error (e.g., α=5% for a 95% CI);

$z_{1-\frac{\alpha}{2}}$ is the standard normal deviate corresponding to a cumulative probability $1-\frac{\alpha}{2}$. E.g., for a symmetrical 95% CI, $\alpha = 5\%$, hence $z_{0.975} \approx 1.96$.

Table 5 Summary of the publisher-specific analysis for the Scopus database

Publisher	d_p	e_p	$\widehat{\varepsilon}_p$	95% Confidence interval (Wilson score interval)		95% Confidence interval (Normal approximation interval)		dev_p
				Lower limit	Upper limit	Lower limit	Upper limit	
(A) Elsevier	2,735	37	0.014	0.010	0.019	0.009	0.018	7.195
(B) MDPI	1,812	22	0.012	0.008	0.018	0.007	0.017	6.714
(C) IEEE	1,648	29	0.018	0.012	0.025	0.011	0.024	0.886
(D) Springer Nature	1,290	59	0.046	0.036	0.059	0.034	0.057	37.817
(E) John Wiley & Sons, Inc	515	18	0.035	0.023	0.055	0.019	0.051	4.817
(F) Taylor & Francis Group	321	8	0.025	0.013	0.049	0.008	0.042	0.241
(G) IOP Publishing Ltd	308	11	0.036	0.021	0.063	0.015	0.056	3.203
(H) ACS Publications	179	1	0.006	0.001	0.031	0.000*	0.017	2.017
(I) SAGE Publications	176	6	0.034	0.017	0.073	0.007	0.061	1.450
(J) American Physical Society (APS)	172	1	0.006	0.001	0.033	0.000*	0.017	1.881
(K) Frontiers Media S.A	141	1	0.007	0.002	0.040	0.000*	0.021	1.293
(L) AIP Publishing LLC	117	1	0.009	0.002	0.048	0.000*	0.025	0.859
(M) Royal Society of Chemistry	104	1	0.010	0.003	0.053	0.000*	0.028	0.638
(N) Association for Computing Machinery, Inc (ACM) Publications	83	1	0.012	0.004	0.067	0.000*	0.036	0.314
(O) Emerald Group Publishing Ltd	76	1	0.013	0.004	0.073	0.000*	0.039	0.220
(P) SPIE	64	0	0.000	0.002	0.058	0.000	0.000	1.341
(Q) Optica Publishing Group	56	3	0.054	0.024	0.152	0.000*	0.113	2.844
(R) American Society of Mechanical Engineers (ASME)	51	1	0.020	0.007	0.107	0.000*	0.058	0.004
(S) Cambridge University Press	45	1	0.022	0.009	0.121	0.000*	0.065	0.003
(T) World Scientific Publishing Co. Pte Ltd	45	0	0.000	0.003	0.082	0.000	0.000	0.943
(U) American Society of Civil Engineers (ASCE)	44	0	0.000	0.004	0.084	0.000	0.000	0.922
(V) Copernicus Publications	42	1	0.024	0.010	0.129	0.000*	0.070	0.016
(W) De Gruyter Brill	40	1	0.025	0.011	0.135	0.000*	0.073	0.031

Table 5 (continued)

Publisher	d_p	e_p	$\widehat{\varepsilon}_p$	95% Confidence interval (Wilson score interval)		95% Confidence interval (Normal approximation interval)		dev_p
				Lower limit	Upper limit	Lower limit	Upper limit	
(X) Oxford University Press	39	2	0.051	0.023	0.178	0.000*	0.121	1.712
(Y) Science Press Ltd	36	0	0.000	0.005	0.102	0.000	0.000	0.754
(Z) Public Library of Science	30	0	0.000	0.007	0.121	0.000	0.000	0.629
Others	665	21	0.032	N/A	N/A	N/A	N/A	N/A
Overall	10,834	227	0.021					78.747

For each p-th publisher (listed in the table rows), the following values are reported: the total number of analysed documents (d_p), the number of observed DT classification errors (e_p), the corresponding (estimated) error rate $\left(\widehat{\varepsilon}_p = \frac{e_p}{d_p}\right)$, the 95% confidence interval (CI) around $\widehat{\varepsilon}_p$ (see Eqs. 6 and 8), and the deviation (dev_p) associated with the χ^2 test (Ross, 2017), which accounts for differences among publishers (see Eqs. 11 and 12). In the third-to-last column, an asterisk (*) indicates cases where the CI lower limit is negative (typically for small values of d_p), which has been replaced with 0. The bottom row reports the overall values for the entire document *corpus*

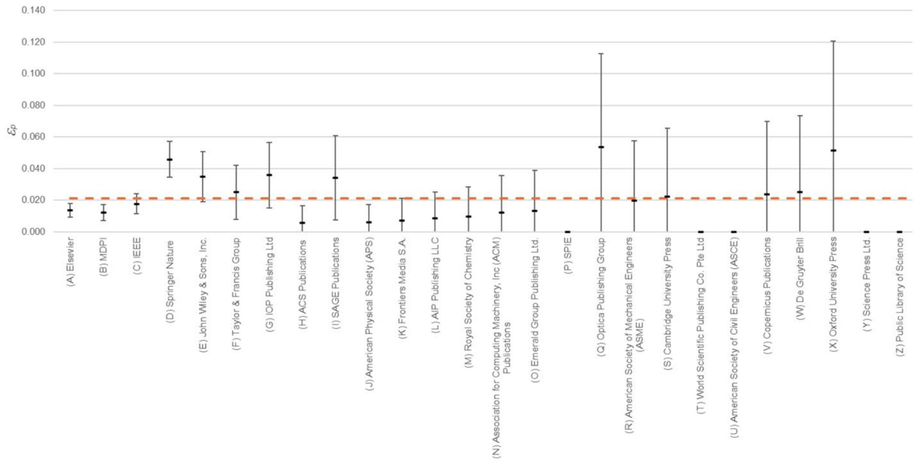


Fig. 4 ε_p values and corresponding 95% CIs (see Eq. 8) for Scopus. The numerical values are provided in Table 5, while publisher abbreviations are listed in Table 4. The horizontal dashed line represents the overall error rate (ε) of the Scopus database

However, when the condition $d_p \cdot \varepsilon_p \geq 5$ is satisfied, E_p can be approximated by a normal distribution with the same mean and variance as above[2]:

$$E_p \sim N\left[\mu = d_p \cdot \varepsilon_p, \sigma^2 = d_p \cdot \varepsilon_p \cdot \left(1 - \varepsilon_p\right)\right] \tag{7}$$

Based on this approximation, a symmetric 95% confidence interval around the estimated error rate $\hat{\varepsilon}_p$ can be defined as follows:

$$\varepsilon_p = \hat{\varepsilon}_p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\varepsilon}_p \cdot \left(1 - \hat{\varepsilon}_p\right)}{d_p}} \tag{8}$$

This approach ensures robust estimates of the ε_p values and their relevant CIs, enabling a statistically rigorous comparison of error rates among publishers. The Wilson score interval is preferred for small sample sizes or extreme proportions, while the normal approximation method offers an alternative when $d_p \cdot \varepsilon_p \geq 5$ (i.e., as in the case for the first seven publishers in Table 5). Comparing both methods provides a more comprehensive view of error rate variability across publishers (Devore, 2016; Ross, 2017). For instance, the $\hat{\varepsilon}_p$ value for documents published by "Springer Nature" is significantly higher than those for "Elsevier", "MDPI" and "IEEE"; this is evident from the non-overlapping 95% CIs around the $\hat{\varepsilon}_p$ values (see also the graphical representation via the boxplot in Fig. 4). Such non-overlap suggests that the "true" error rates for these publishers are statistically distinct, implying systematic differences in how documents from different publishers are classified.

The *chi-squared* (χ^2) test confirms these considerations, evaluating whether the observed error rates differ significantly across publishers. For each p-th publisher, the

[2] This condition is generally met when considering relatively large sets of (misclassified) documents (e_p). However, even when it is not strictly met, Eq. 8 still provides a practical and reasonable estimate of the CI (Ross, 2017).

deviation dev_p, representing the standardized residuals of the χ^2 test, is calculated as follows (Ross, 2017):

$$dev_p = \frac{\left(O_p - E_p\right)^2}{E_p}, \tag{9}$$

where $O_p = e_p = \varepsilon_p \cdot d_p$ is the *observed* number of errors and $E_p = \varepsilon \cdot d_p$ is the *expected* number of errors under the null hypothesis (H_0) of equal error rates across publishers.

The deviation relative to the p-th publisher therefore becomes:

$$dev_p = \frac{\left(\varepsilon_p \cdot d_p - \varepsilon \cdot d_p\right)^2}{\varepsilon \cdot d_p} = \frac{\left(\varepsilon_p - \varepsilon\right)^2}{\varepsilon} \cdot d_p \tag{10}$$

The test statistic χ^2 is then computed as:

$$\chi^2 = \sum_{\forall p} dev_p = \sum_{\forall p} \frac{\left(\varepsilon_p - \varepsilon\right)^2}{\varepsilon} \cdot d_p, \tag{11}$$

which is compared against the critical value $\chi^2_{\alpha,DoF}$ for the chosen significance level (e.g., $\alpha = 5\%$) and *degrees of freedom* (*DoF*), determined as:

$$DoF = (rows - 1) \cdot (colums - 1), \tag{12}$$

as a function of the number of rows and columns of the table containing the data of interest. In this case, $rows = 2$ (distinguishing between correctly classified and misclassified documents), while the *columns* parameter is equal to the number of publishers, determining $DoF = (2-1) \cdot (26-1) = 25$.

The comparison between the calculated χ^2 value and the critical one leads to two possible outcomes:

$$\begin{aligned} &\chi^2 > \chi^2_{\alpha,DoF} (\text{or } p < \alpha) : \text{ Reject the null hypothesis } \left(\mathrm{H}_0\right); \\ &\chi^2 \leq \chi^2_{\alpha,DoF} (\text{or } p \geq \alpha) : \text{ Do not reject } \mathrm{H}_0, \end{aligned} \tag{13}$$

where H_0 is that there are no significant differences between publishers in terms of DT-classification error rates.

In the specific case under analysis, the result is:

$$\chi^2 = 78.747 > \chi^2_{\alpha,DoF} = 37.652\left(p = 2 \times 10^{-7} < \alpha = 0.05\right), \tag{14}$$

which leads to rejecting H_0.

From the dev_p values (cf. Table 5), it becomes evident that certain publishers deviate more significantly than others. For instance, "Springer Nature", "IOP Publishing Ltd" and "John Wiley & Sons, Inc." have significantly higher ε_p values, while "Elsevier", "MDPI" and "IEEE" exhibit significantly lower values compared to the rest.

Table 6 includes an additional test on pairwise differences in error rates. Specifically, for each pair of publishers (e.g., $p = A$ and B), the null hypothesis (H_0) assumes that the two error rates (ε_p) are equal, while the alternative hypothesis (H_1) assumes that they differ. The test statistic z is defined as follows:

Table 6 Contingency table of $|z|$-values for pairwise comparisons of error rates

	(A) Elsevier	(B) MDPI	(C) IEEE	(D) Springer Nature	(E) John Wiley & Sons, Inc	(F) Taylor & Francis Group
(A) Elsevier	–					
(B) MDPI	0.40	–				
(C) IEEE	1.07	1.33	–			
(D) Springer Nature	**6.25**	**5.78**	**4.44**	–		
(E) John Wiley & Sons, Inc	**3.46**	**3.51**	**2.36**	1.02	–	
(F) Taylor & Francis Group	1.60	1.79	0.88	1.67	0.81	–
(G) IOP Publishing Ltd	**2.96**	**3.09**	**2.06**	0.77	0.06	0.79

Values where $|z| > 1.96$ are highlighted in bold

$$z = \frac{\varepsilon_A - \varepsilon_B}{\sqrt{\varepsilon_{pool} \cdot \left(1 - \varepsilon_{pool}\right) \cdot \left(\frac{1}{d_A} + \frac{1}{d_B}\right)}}, \tag{15}$$

where:

ε_A and ε_B are the observed error rates for the two publishers of interest, calculated as: $\varepsilon_A = \frac{e_A}{d_A}$, $\varepsilon_B = \frac{e_B}{d_B}$;

d_A and d_B are the numbers of documents analysed for each publisher;

ε_{pool} is the pooled error rate, computed as: $\varepsilon_{pool} = \frac{e_A + e_B}{d_A + d_B}$.

In practice, after computing the z-statistic (Eq. 15), it is compared with the critical value $z_{1-\frac{\alpha}{2}} \approx 1.96$ for $a = 5\%$. If $|z| > 1.96$, the null hypothesis is rejected, indicating that the two error rates (ε_A and ε_B) are statistically different (Ross, 2017).

For simplicity, the analysis was limited to the top 7 largest publishers in terms of the number of indexed (and also misclassified) documents (cf. Table 5). Among all the $\binom{7}{2} = \frac{7 \cdot (7-1)}{2} = 21$ possible pairwise comparisons, three show statistically significant differences, confirming the findings of the previous statistical tests. The pairwise comparisons revealing the largest differences involve publishers "Springer Nature", "John Wiley & Sons, Inc." and "IOP Publishing Ltd". More details on the distribution of DT-classification errors by Scopus from the perspective of each publisher are provided in online Appendix A.1.

Publisher-specific analysis for WoS

Table 7, similar to Table 5 (for Scopus), summarizes the results of the analysis for WoS. As with Scopus (cf. sect. "Publisher-specific analysis for Scopus"), various error statistics were determined for each of the most relevant publishers. The χ^2 test does not indicate significant differences among publishers:

$$\chi^2 = 21.022 \leq \chi^2_{\alpha,DoF} = 37.652\,(p = 0.69 \geq \alpha = 0.05), \tag{16}$$

Table 7 Summary of the publisher-specific analysis for the WoS database

Publisher	d_p	e_p	$\hat{\varepsilon}_p$	95% Confidence interval (Wilson score interval)		95% Confidence interval (Normal approximation interval)		dev_p
				Lower limit	Upper limit	Lower limit	Upper limit	
(A) Elsevier	2,735	32	0.012	0.008	0.016	0.008	0.016	0.231
(B) MDPI	1,812	21	0.012	0.008	0.018	0.007	0.017	0.188
(C) IEEE	1,648	15	0.009	0.006	0.015	0.005	0.014	1.710
(D) Springer Nature	1,290	19	0.015	0.009	0.023	0.008	0.021	0.401
(E) John Wiley & Sons, Inc	515	5	0.010	0.004	0.023	0.001	0.018	0.371
(F) Taylor & Francis Group	321	8	0.025	0.013	0.049	0.008	0.042	3.741
(G) IOP Publishing Ltd	308	8	0.026	0.014	0.051	0.008	0.044	4.236
(H) ACS Publications	179	2	0.011	0.004	0.040	0.000*	0.027	0.034
(I) SAGE Publications	176	3	0.017	0.006	0.050	0.000*	0.036	0.256
(J) American Physical Society (APS)	172	1	0.006	0.001	0.033	0.000*	0.017	0.647
(K) Frontiers Media S.A	141	2	0.014	0.005	0.051	0.000*	0.034	0.023
(L) AIP Publishing LLC	117	1	0.009	0.002	0.048	0.000*	0.025	0.161
(M) Royal Society of Chemistry	104	0	0.000	0.001	0.036	0.000	0.000	1.325
(N) Association for Computing Machinery, Inc (ACM) Publications	83	0	0.000	0.001	0.045	0.000	0.000	1.057
(O) Emerald Group Publishing Ltd	76	2	0.026	0.010	0.093	0.000*	0.062	1.100
(P) SPIE	64	0	0.000	0.002	0.058	0.000	0.000	0.815
(Q) Optica Publishing Group	56	1	0.018	0.007	0.098	0.000*	0.053	0.115
(R) American Society of Mechanical Engineers (ASME)	51	0	0.000	0.003	0.073	0.000	0.000	0.650
(S) Cambridge University Press	45	0	0.000	0.003	0.082	0.000	0.000	0.573
(T) World Scientific Publishing Co. Pte Ltd	45	0	0.000	0.003	0.082	0.000	0.000	0.573
(U) American Society of Civil Engineers (ASCE)	44	0	0.000	0.004	0.084	0.000	0.000	0.560
(V) Copernicus Publications	42	1	0.024	0.010	0.129	0.000*	0.070	0.404
(W) De Gruyter Brill	40	0	0.000	0.004	0.092	0.000	0.000	0.510

Table 7 (continued)

Publisher	d_p	e_p	$\hat{\varepsilon}_p$	95% Confidence interval (Wilson score interval)		95% Confidence interval (Normal approximation interval)		dev_p
				Lower limit	Upper limit	Lower limit	Upper limit	
(X) Oxford University Press	39	0	0.000	0.004	0.094	0.000	0.000	0.497
(Y) Science Press Ltd	36	0	0.000	0.005	0.102	0.000	0.000	0.459
(Z) Public Library of Science	30	0	0.000	0.007	0.121	0.000	0.000	0.382
Others	665	17	0.026	N/A	N/A	N/A	N/A	N/A
Overall	10,834	138	0.013					21.022

For each p-th publisher (listed in the table rows), the following values are reported: the total number of analysed documents (d_p), the number of observed DT classification errors (e_p), the corresponding (estimated) error rate $\left(\hat{\varepsilon}_p = \frac{e_p}{d_p}\right)$, the 95% confidence interval (CI) around $\hat{\varepsilon}_p$ (see Eqs. 6 and 8), and the deviation (dev_p) associated with the χ^2 test, which accounts for differences among publishers (see Eqs. 11 and 12). In the third-to-last column, an asterisk (*) indicates cases where the CI lower limit is negative (typically for small values of d_p), which has been replaced with 0. The bottom row reports the overall values for the entire document *corpus* which leads to failing to reject H_0, meaning that there is no strong statistical evidence of significant differences in DT-classification error rates among publishers. This result is further supported by the diagram in Fig. 5, which shows that the CIs around the ε_p values for different publishers tend to overlap, albeit slightly. However, "Taylor & Francis Group" and "IOP Publishing Ltd" appear to exhibit higher error rates than the rest, despite their CIs still overlapping with those of other publishers. This observation is further confirmed by the pairwise-comparison test of ε_p values, which is restricted to the seven largest publishers in terms of indexed documents (cf. Table 8). The test highlights some significant pairwise differences, notably involving “Taylor & Francis Group” and “IOP Publishing Ltd”.

In summary, while the χ^2 test does not reveal significant differences in error rates across publishers, the test on pairwise differences in ε_p values reveals statistically significant differences. Therefore, we can conclude that some systematic differences among publishers seem to exist in WoS as well, albeit to a lesser extent than in Scopus. More details on the distribution of DT-classification errors by WoS from the perspective of each publisher are provided in online Appendix A.2.

Comparison between Scopus and WoS

The results from each database’s perspective can be summarized into the map in Fig. 6, which plots the ε_p values for Scopus on the x-axis and those for WoS on the y-axis, for each of the 26 macro-publishers. Notably, there appears to be no significant correlation between the publishers that are most critical for one database, in terms of ε_p values, and those that are most critical for the other, as indicated by the relatively low $R^2 \approx 0.16$. This finding suggests that the two databases may differ in their familiarity with handling DTs for specific publishers. The causes of this apparent independence will be investigated in future studies, exploring various possible hypotheses, including:

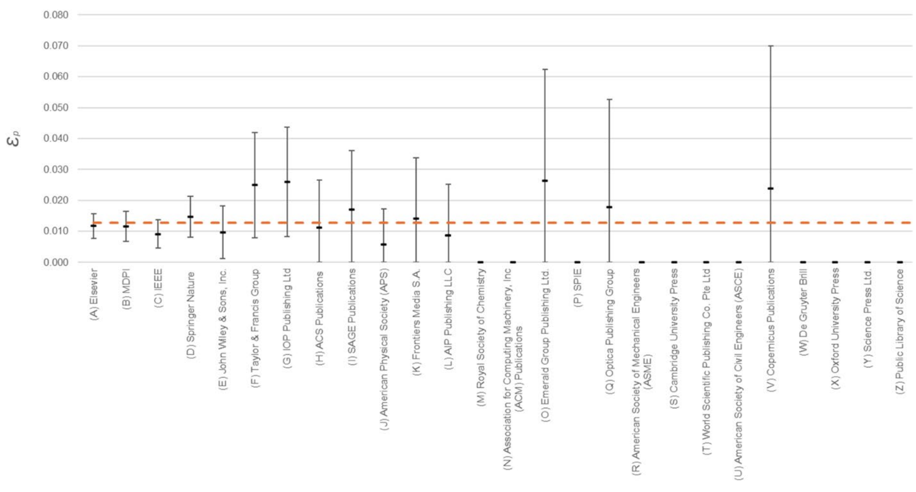
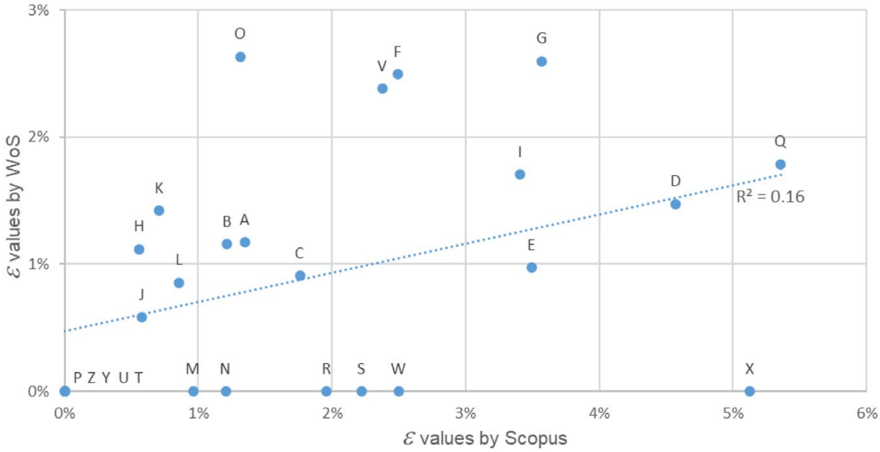


Fig. 5 ε_p values and corresponding 95% CIs (see Eq. 8) for WoS. The numerical values are provided in Table 7, while publisher abbreviations are listed in Table 4. The horizontal dashed line represents the overall error rate (ε) of the WoS database

Table 8 Contingency table of $|z|$-values for pairwise comparisons of error rates

	(A) Elsevier	(B) MDPI	(C) IEEE	(D) Springer Nature	(E) John Wiley & Sons, Inc	(F) Taylor & Francis Group	(G) IOP Publishing Ltd
(A) Elsevier	–						
(B) MDPI	0.03	–					
(C) IEEE	0.81	0.72	–				
(D) Springer Nature	0.80	0.76	1.42	–			
(E) John Wiley & Sons, Inc	0.39	0.36	0.13	0.84	–		
(F) Taylor & Francis Group	**1.97**	1.90	**2.41**	1.27	1.73	–	
(G) IOP Publishing Ltd	**2.09**	**2.01**	**2.52**	1.38	1.81	0.08	–

Values where $|z| > 1.96$ are highlighted in bold



Macro-publisher		Scopus	WoS
(A)	Elsevier	1.4%	1.2%
(B)	MDPI	1.2%	1.2%
(C)	IEEE	1.8%	0.9%
(D)	Springer Nature	4.6%	1.5%
(E)	John Wiley & Sons, Inc.	3.5%	1.0%
(F)	Taylor & Francis Group	2.5%	2.5%
(G)	IOP Publishing Ltd	3.6%	2.6%
(H)	ACS Publications	0.6%	1.1%
(I)	SAGE Publications	3.4%	1.7%
(J)	American Physical Society (APS)	0.6%	0.6%
(K)	Frontiers Media S.A.	0.7%	1.4%
(L)	AIP Publishing LLC	0.9%	0.9%
(M)	Royal Society of Chemistry	1.0%	0.0%
(N)	Association for Computing Machinery, Inc (ACM) Publications	1.2%	0.0%
(O)	Emerald Group Publishing Ltd.	1.3%	2.6%
(P)	SPIE	0.0%	0.0%
(Q)	Optica Publishing Group	5.4%	1.8%
(R)	American Society of Mechanical Engineers (ASME)	2.0%	0.0%
(S)	Cambridge University Press	2.2%	0.0%
(T)	World Scientific Publishing Co. Pte Ltd	0.0%	0.0%
(U)	American Society of Civil Engineers (ASCE)	0.0%	0.0%
(V)	Copernicus Publications	2.4%	2.4%
(W)	De Gruyter Brill	2.5%	0.0%
(X)	Oxford University Press	5.1%	0.0%
(Y)	Science Press Ltd.	0.0%	0.0%
(Z)	Public Library of Science	0.0%	0.0%

Fig. 6 Map comparing the ε_p results for the two databases: Scopus (x-axis) and WoS (y-axis). Further publisher-specific statistics are provided in online Appendix A.1 (for Scopus) and online Appendix A.2 (for WoS)

- The two databases may adopt different DT classification rules, which could align more or less closely with the editorial standards of specific publishers, leading to varying error patterns.
- Some publishers may have a closer relationship with a given database. For instance, Elsevier and the Scopus database are both part of the same organization (RELX Group),

which could reduce the likelihood of classification errors or facilitate their prompt correction when they occur (Franceschini et al., 2016a).

Concluding remarks

This research highlighted several findings of interest for academic researchers. Firstly, while DT-classification errors can be considered relatively rare events, their occurrence is nonetheless significant in both Scopus (error rate ~ 2.1%) and WoS (error rate ~ 1.3%). A deeper analysis from the publishers' perspective revealed statistically significant differences in their propensity to incur DT-classification errors. Specifically, Scopus appears to struggle more with publishers "Springer Nature", "IOP Publishing Ltd" and "John Wiley & Sons, Inc.", with relevant e_p values of ~ 4.6%, ~ 3.6% and ~ 3.5% respectively, whereas WoS shows higher error rates with publishers "IOP Publishing Ltd" and "Taylor & Francis Group", with relevant e_p values of ~ 2.6% and ~ 2.5% respectively (cf. RQ1).

The reasons behind these discrepancies remain to be thoroughly investigated, but they could include various factors (cf. RQ2). For example, some editorial styles or conventions in DT-classification may align more closely with the internal classification rules of one database than the other. Additionally, certain publishers may predominantly include documents with more "canonical" or common DTs, which are inherently easier to classify accurately. Moreover, larger publishers span a broader range of disciplines—even within the same field—and manage extensive portfolios of journals and books; this increased diversity of DTs may lead to a higher likelihood of DT-classification errors. Conversely, smaller publishers – which often focused on a limited number of well-established DTs – are likely to exhibit a lower propensity for DT-classification errors. Another contributing factor could be the impact of mergers and acquisitions (M&A), where smaller publishers are absorbed by larger groups. In such cases, editorial standards may either remain unchanged or be adjusted to align with those of the acquiring group, potentially affecting DT-classification consistency.

From a practical standpoint, this work aims to raise awareness of the possibility of DT-classification errors and the fact that, depending on the database in use (e.g., Scopus or WoS), documents from certain publishers may be at a higher risk of misclassification. Therefore, researchers—particularly those involved in competitive evaluations for promotions or scientific assessments—are advised to carefully verify not only the correct indexing of documents but also the accuracy of their assigned DTs. In general, although Scopus and WoS do not seem to routinely conduct periodic reviews of the accuracy of indexed data, such as by cross-referencing results with those of competing databases, they are relatively responsive in addressing reported errors through designated channels (Franceschini et al., 2016b).

This analysis has some limitations, such as being based on a sample of documents that, while extensive, is confined to the field of *engineering*. Consequently, the results may not be immediately generalizable to other fields (Franceschini & Maisano, 2011; Franceschini et al., 2012). Furthermore, some documents were excluded from the analysis, such as those lacking DOI codes or those with multiple DT-classifications assigned in WoS.[3]

[3] Unlike Scopus, WoS permits multiple classifications for a significant portion of documents (approximately 25%), often applying two DT categories simultaneously (e.g., *article + data paper*, *review + book chapter*, etc.) (Maisano et al., 2025b). For simplicity, DTs in WoS with multiple classifications were excluded from this analysis.

Regarding the future, it is planned to expand the analysed sample to include a broader range of scientific fields and incorporate documents with multiple DT-classifications. Additionally, the specific causes underlying the systematic differences between publishers observed in this study will be explored in greater depth.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-025-05444-6.

Acknowledgements This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and was partially funded by the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Author contributions The authors have provided an equal contribution to the drafting of the paper.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. MICS (Made in Italy – Circular and Sustainable) Extended Partnership and was partially funded by the European Union Next-GenerationEU, PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004.

Data availability The original data cannot be provided as a subscription is required to access the databases used.

Declarations

Conflict of interest The authors have no conflict of interest.

Ethical approval The authors respect the Ethical Guidelines of the journal.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Baas, B., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies, 1*(1), 377–386.

Cambridge University Press. (2023). Academic publishing guides: Editorial services style guide for academic books, version 5.0. Retrieved February, 2025, from https://www.cambridgescholars.com/uploads/sbpub_sub/downloads/Author-Guidance.pdf

Clarivate. (2025). Web of science help—Document types. Retrieved February, 2025, from https://webofscience.help.clarivate.com/en-us/Content/document-types.html

Crotty, D. (2024). Variability, irregular publisher metadata, and the ongoing evolution of databases complicates reproducibility in bibliometrics research. Society for Scholarly Publishing. The Scholarly Kitchen blog. Retrieved February, 2025, from https://scholarlykitchen.sspnet.org/2024/08/15/varia bility-bad-publisher-metadata-and-the-ongoing-evolution-of-databases-makes-bibliometrics-research-reproducibility-difficult/

Csomós, G., & Farkas, J. Z. (2023). Understanding the increasing market share of the academic publisher "Multidisciplinary Digital Publishing Institute" in the publication output of Central and Eastern European countries: A case study of Hungary. *Scientometrics, 128*, 803–824.

Devore, J. L. (2016). *Probability and statistics for engineering and the sciences* (9th ed.). Cengage.

Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics, 113*(1), 219–236.

Donner, P. (2023). Data inaccuracy quantification and uncertainty propagation for bibliometric indicators. arXiv:2303.16613

Elango, B. (2025). Assigning different document types by Scopus for similar contents: An exploratory analysis. *Scientometrics*. https://doi.org/10.1007/s11192-025-05270-w

Elsevier. (2025). Scopus content coverage guide. Retrieved February, 2025, from https://www.elsevier.com/products/scopus/content

Franceschini, F., & Maisano, D. (2011). Bibliometric positioning of scientific manufacturing journals: A comparative analysis. *Scientometrics, 86*(2), 463–485.

Franceschini, F., & Maisano, D. (2014). Sub-field normalization of the IEEE scientific journals based on their connection with technical societies. *Journal of Informetrics, 8*(3), 508–533.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2013). A novel approach for estimating the omitted citation rate of bibliometric databases with an application to the field of bibliometrics. *Journal of the American Society for Information Science and Technology, 64*(10), 2149–2156.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2014). Scientific journal publishers and omitted citations in bibliometric databases: Any relationship? *Journal of Informetrics, 8*(3), 751–765.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. *Scientometrics, 102*(3), 2181–2186.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016a). Empirical analysis and classification of database errors in Scopus and Web of Science. *Journal of Informetrics, 10*(4), 933–953.

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016b). Do Scopus and WoS correct "old" omitted citations? *Scientometrics, 107*, 321–335.

Franceschini, F., Maisano, D., & Turina, E. (2012). European research in the field of production technology and manufacturing systems: An exploratory analysis through publications and patents. *The International Journal of Advanced Manufacturing Technology, 62*(1–4), 329–350.

García-Pérez, M. A. (2010). Web of Science, PsycINFO, and Google scholar: A case study for the computation of h indices in psychology. *Journal of the American Society for Information Science and Technology, 61*(10), 2070–2085.

Harzing, A. W. (2013). Document categories in the ISI Web of Knowledge: Misunderstanding the social sciences? *Scientometrics, 94*(1), 23–34.

Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., & Mayr, P. (2024). Analysis of the publication and document types in OpenAlex, Web of Science, Scopus, PubMed and Semantic Scholar. arXiv:2406.15154

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE, 10*(6), Article e0127502.

Maisano, D.A., Ferrara, L., Franceschini, F. (2025b). Impact of Web of Science and Scopus Policies on multiple document-type classification. In *Proceedings of the 20th international conference on scientometrics & informetrics (ISSI 2025)* (pp. 968–984), 23–27 June 2025, Yerevan (Armenia)

Maisano, D. A., Mastrogiacomo, L., Ferrara, L., & Franceschini, F. (2025a). A large-scale semi-automated approach for assessing document-type classification errors in bibliometric databases. *Scientometrics, 130*, 1901–1938.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer. https://doi.org/10.1007/1-4020-3714-7

Mokhnacheva, Y. V. (2023). Document types indexed in WoS and Scopus: Similarities, differences, and their significance in the analysis of publication activity. *Scientific and Technical Information Processing, 50*(1), 40–46.

Olensky, M., Schmidt, M., & Van Eck, N. J. (2016). Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of Science. *Journal of the Association for Information Science and Technology, 67*(10), 2550–2564.

Ross, S. M. (2017). *Introductory statistics*. Academic Press (Elsevier).

Sigogneau, A. (2000). An analysis of document types published in journals related to physics: Proceeding papers recorded in the science citation index database. *Scientometrics, 47*(3), 589–604.

Valderrama-Zurián, J.-C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics, 9*(3), 570–576.
van Bellen, S., Alperin, J. P., Larivière, V. (2024). The oligopoly of academic publishers persists in exclusive database. arXiv:2406.17893
Yeung, A. W. K. (2021). Document type assignment by Web of Science, Scopus, PubMed, and publishers to "Top 100" papers. *Malaysian Journal of Library & Information Science, 26*(3), 97–103.
Yeung, A. W. K. (2019). Comparison between Scopus, Web of Science, PubMed and publishers for mislabelled review papers. *Current Science, 116*(11), 1909–1914.
Zhu, M., Lu, X., Chen, F., Yang, L., & Shen, Z. (2024). An explorative study on document type assignment of review articles in web of science, Scopus and journals' websites. *Journal of Data and Information Science, 9*(1), 11–36.