Bulletin of the *Transilvania* University of Brasov
Series II: Forestry • Wood Industry • Agricultural Food Engineering • Vol. 16(65) No. 1 – 2023
https://doi.org/10.31926/but.fwiafe.2023.16.65.1.5

SCIENTIFIC METRICS ANALYSIS OF LITERATURE RELATED TO SUSTAINABLE WOOD PROJECTS

Adrienn NOVOTNI[1] **Zsolt TÓTH[1]**

Abstract: *An analysis of the literature on wood projects (n = 205) was conducted to understand current scientific trends, key issues, and the appropriate representation of scientific results. Four scientific questions were addressed: whether the number of citations in the literature follows a power function distribution after a value, whether there is any relationship between the number of articles and the number of citations per year, whether the publications available in Scopus can be considered a "true" representation of the publications, and whether the Lotka distribution for author productivity holds for the publication sample on wood projects. The power function distribution test shows that most articles have very few citations, while some articles have very many. This suggests that the scientific value of articles is not necessarily determined by the number of citations. There is a correlation between the number of publications and citations over the same period or slightly skewed in time. The Bradford distribution suggests that publications available in Scopus describing wood projects cannot be considered as a real representation of the publications on the subject. Despite this, the frequency of author productivity follows a Lotka distribution.*

Key words: *sustainable wood, wood projects, scientific metrics, distribution analysis, citation analysis.*

1. Introduction

An environmentally realistic view of highly processed wood products, better marketing, making the whole product life cycle more sustainable, and improving the profitability of “forest” and “wood” at the local community level can be key factors in preserving forests and meeting sustainability targets. The increasing high-tech use of wood can reinforce positive trends. However, this requires an appropriate level of financial resources, which in recent decades has been complemented by a project approach, i.e. addressing specific objectives through projects. A review of the literature on the subject, analysed using scientific metrics and text-mining methods, can contribute to an understanding of current scientific trends, key issues, and the appropriate representation of scientific results.

Four scientific questions were addressed in the analysis using statistical tools.

[1] University of Sopron, Faculty of Wood Engineering and Creative Industries, Sopron, Hungary; Correspondence: Tóth Zsolt; email: toth.zsolt@uni-sopron.hu.

1) The first problem investigated was whether the number of citations in the literature on wood projects describes a power function distribution after a certainx_{min}, like many other economic and social phenomena [1]. This essentially corresponds to the question of "Do most articles get very few citations, while some articles get very many?".
2) The second question was whether there was any relationship between the number of articles and the number of citations per year for wood projects. Possible time lags were also taken into account here.
3) The third research question examined is whether the publications available in Scopus describing wood projects can be considered a relatively "true" representation of the publications on the subject.
4) Fourth, the Lotka distribution for author productivity was examined to see whether it is valid for the publication sample on wood projects.

The research methods used by the authors have, of course, a long history in the literature, but relatively few of them are strictly related to wood projects. Nevertheless, from a methodological and thematic conceptual point of view, there are undoubtedly noteworthy reviews related to the methods of modelling in Industry 4.0 wood research [9], to the subject of bioeconomy hijacked by public institutions to announce economic and ecological transition [15], to Responsible Sourcing (RS) Program supporting the Natural Resource Justice (NRJ) in energy transitions, based on a bibliometric network analysis, to the trade specializations in wood of a selected country [8], or to the critical factors that stand as a barrier between manufacturing and environmental sustainability in wood furniture industries [6]. Other papers are rather tangential in their treatment of the subject [3], [5], [10], [12].

2. Materials and Methods

2.1. General Information

A BibTeX file describing articles related to wood projects, extracted from Scopus, was loaded into a bibliographic data frame and bibliographic statistics were generated. In addition to the basic R base packages, mainly the R bibliometrix [2] package was used extensively in the analysis.

2.2. Power Function Distribution of Citations in the Literature on Wood Projects

The parameters *γ* and *ζ* of the following probability distribution are sought (Eq. (1)):

$$p(x) = \frac{x^{-\gamma}}{\xi(\alpha, \gamma_{min})} \quad (1)$$

where *n* is an empirical value (Eq. (1)):

$$\xi(\gamma, x_{min}) = \sum_{n=0}^{\infty} (n + x_{min})^{-\gamma} \quad (2)$$

Thus, a power function with parameter γ is looked for, which also describes the empirical values from $x_{min} > 0$. The hypothesis test is based on the Kolmogorov-Smirnov test (α = 0.05, for all the tests in the paper):

H_0: The distribution is derived from a power function distribution.

Due to the extreme values, a bootstrapping method with 1,000 iterations was used following the literature advice [4].

2.3. Relationship between the Number of Articles and the Number of Citations

Correlation tests were used to examine research question 2. Because of the presumed lack of normality (tested with Kolmogorov-Smirnov test and Shapiro-Wilk test) Spearman's rank correlation coefficient was calculated.

$$H_o : \rho = 0 \tag{3}$$

2.4. Distribution of Papers in Journals

Research question 3 was examined using Bradford's law. According to the version of Bradford's law for journals and articles [11], if we sort journals in descending order of the number of articles they contain and then create three thirds – also following the descending order – where the number of articles in each third is the same, then the ratio of the number of journals in each third starting from the top third is $1:n:n^2$, where n is the Bradford value (or Bradford constant). However, the original procedure was, and still is, valid for referencing the holdings of large libraries, grouped according to different criteria, without a sound scientific explanation. For smaller samples, a common method is to use the Leimkuhler model based on Bradford's law [14].

For three zones, assuming a Bradford distribution, the Equation (4):

$$k = \left(1.781 y_m\right)^{\frac{1}{3}} \tag{4}$$

where:

k is the Bradford constant and

y_m is the number of articles in the journal with the most articles.

The coefficient of the Bradford constant can be calculated by the Equation (5):

$$r_o = \frac{T\left(k-1\right)}{k^3 - 1} \tag{5}$$

where T is the number of journals.

If the empirical distribution differs from the Bradford distribution, we can assume that the sample under investigation is biased for some reason. There is no widely used exact statistical estimation procedure for the Bradford distribution.

2.5. Lotka Distribution

A statistical hypothesis that is equivalent to research question 4 was tested, namely:

$$H_o: y \approx c/x^{\beta} \Delta\beta \approx 2 \tag{6}$$

where:

y is the number of authors of x articles and

c is a constant.

To compare the empirical distribution with the theoretical Lotka distribution, a two-sample Kolmogorov-Smirnov test was performed.

It may be worth noting that the sum of calculated theoretical values often gives a value greater than 1 [13].

3. Results

3.1. General Results

Table 1 shows the key data from the descriptive statistics of the publications examined.

Table 1

Key data on publications related to wood projects

Indicator	Value
Years	1984-2022
Number of documents	205 units
Average annual growth	6.48%
Average number of citations	8.37
Annual average	0.74
Number of sources cited	4,198 units
Average age of publications (19.10.2022)	11.9 years

3.2. Results of Power Function Distribution of Citations

Figure 1 seems to visually confirm the power function distribution.

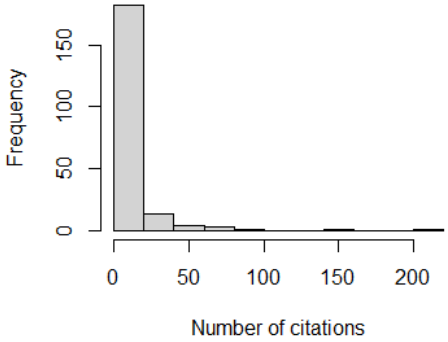

Fig. 1. *Number of citations of publications*

With a bootstrapping procedure based on 1000 iterations, optimised to estimate p-value and to test the distribution, x_{min}=12.59, p=0.61, γ=2.23. Since p>α and 2<γ<3, $H_0(3)$ for the power function distribution is not rejected (Figure 2).

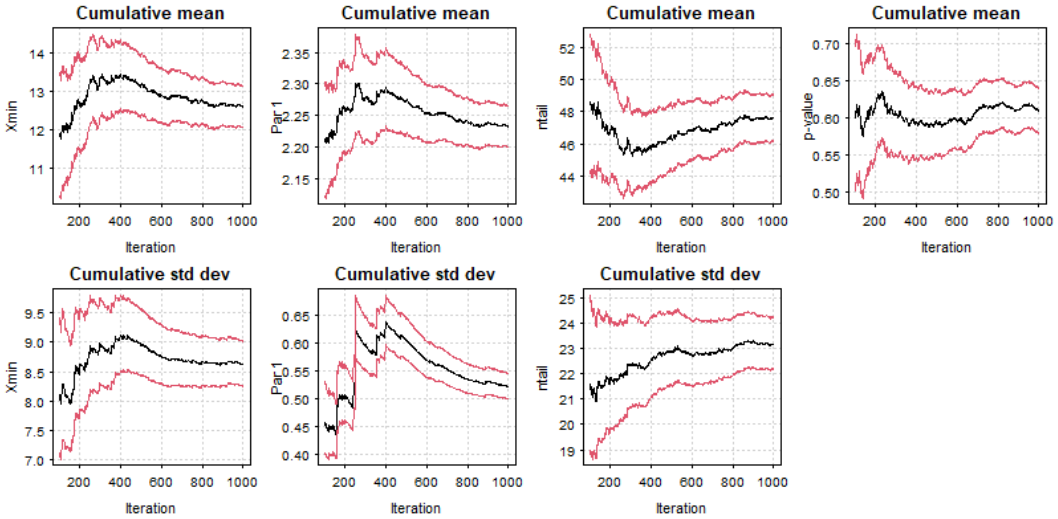

Fig. 2. *Number of citations of publications*

However, it should also be stressed that for values smaller than x_{min} the power function distribution does not hold. Out of 205 related articles, on average only 48 papers fall in the range larger than x_{min} during the iterations. Moreover, only 32 values are higher than the average x_{min} obtained after 1,000 iterations. A significant number of articles (91) have no citation.

3.3. Correlation Test between the Number of Articles and the Number of Citations

The p-value of the Kolmogorov-Smirnov test and Shapiro-Wilk test for normality was well below the significance level for both variables. Due to the lack of normality, Pearson's correlation cannot be calculated, although it is possible to calculate Spearman's rank correlation.

H_0(4) for independence was rejected not only for publications and citations in the same year, but also for citations one or two years later than the publications since $p<\alpha$ holds here as well. Thus, there is a medium correlation between publications and citations in years 0-2. Even in the later years, ρ was occasionally higher than the second year, but p was higher than α in all cases, so that H_0 could not be clearly rejected (Table 2). The perhaps unconventional delimitation of the years of publications and citations was justified on the one hand by the need to ensure a possible time lag and on the other hand by the equal number of elements required for correlation analysis of the two time series under study.

3.4. Analysis of Bradford Distribution

The values of Table 3 reflect a high number of journals relatively to the number of papers.

Table 2

Spearman's rank correlation test for number of publications and citations

Years of publications	Years of citations	ρ	p	Decision
1984-2022	1984-2022	0.52	0.0007	H_0 is rejected.
1984-2021	1985-2022	0.42	0.008	H_0 is rejected.
1984-2020	1986-2022	0.45	0.005	H_0 is rejected.
1984-2019	1987-2022	0.25	0.14	H_0 is not rejected.

Table 3

Bradford zones

Zone	Number of journals	Distribution	Number of papers
1.	24	15.4%	69
2.	65	41.7%	69
3.	67	42.9%	67
Σ	156	100%	205

The ratio of the number of journals in each zone is not close to the Bradford $1:n:n^2$ empirical values. k = 2.42 according to (4), r_0 = 16.77 according to (5). The

Bradford distribution calculated by Egghe's formula [7] is then 16.77:40.66:98.57 in the three thirds. It can be seen that our empirical values do not follow a Bradford distribution.

3.5. Analysis of Lotka Distribution

The β of the two-sample Kolmogorov-Smirnov test of Lotka's distribution (4.72) differs significantly from the general value originally estimated by Lotka (β=2). However, this does not necessarily mean that the frequency distribution of author productivity does not follow Lotka's distribution. The p-value (0.52) is significantly higher than the significance level. Thus, $H_0(6)$ is not definitely rejected (Figure 3).

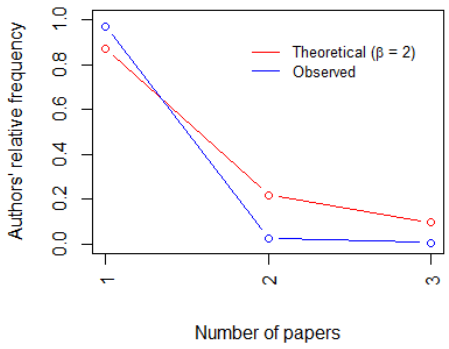

Fig. 3. *Theoretical and observed Lotka distribution*

4. Discussion

The power function distribution test shows that most articles have very few citations, while some articles have very many. This phenomenon can be assumed as general. The relevant question is whether this really reflects the professional value of the articles. There are many articles that are hardly cited or not cited at all. However, the exact citation rate of the less cited articles - often old, outdated or of less interest to the scientific community in the most developed countries for various reasons - cannot necessarily be identified from Scopus, and may be higher when taking into account articles not included in Scopus. However, this is very difficult to verify.

Although the method is not perfect, there is a correlation between the number of publications and citations over the same period or slightly skewed in time. It is assumed that citation rates depend to a significant extent on the number of publications.

Based on the analysis related to the Bradford distribution, the publications available in Scopus describing wood projects cannot be considered as a real representation of the publications on the subject.

Although the result is not completely straightforward, we do not reject that the frequency of author productivity follows a Lotka distribution.

5. Conclusions

Based on a scientific metrics analysis of publications related to timber projects, the publication strategy for future research should not be pinned down to any particular strategy. It is certainly worth targeting the highest-ranked journals, but the majority of articles will be low-cited in any case. It is likely that a lively network of contacts through projects and the scientific values of outputs will guarantee high citation rates, but it is not necessarily true that less cited articles do not carry serious scientific content.

References

1. Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. In: Reviews of Modern Physics, vol. 74(1), pp. 47-97. DOI: 10.1103/RevModPhys.74.47.
2. Aria, M., Cuccurullo, C., 2017. Bibliometrix: An R-tool for comprehensive science mapping analysis. In: Journal of Informetrics, vol. 11(4), pp. 959-75. DOI: 10.1016/j.joi.2017.08.007.
3. Callegari, B., Nybakk, E., 2022. Schumpeterian theory and research on forestry innovation and entrepreneurship: The state of the art, issues and an agenda. In: Forest Policy and Economics, vol. 138, ID article 102720. DOI: 10.1016/j.forpol.2022.102720.
4. Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. In: SIAM Review, vol. 51(4), pp. 661-703. DOI: 10.1137/070710111.
5. Cong, W., Shi, L., 2019. Heterogeneity of industrial development and evolution of cleaner production: Bibliometric analysis based on JCLP. In: Journal of Cleaner Production, vol. 212, pp. 822-836. DOI: 10.1016/j.jclepro.2018.12.024.
6. Dalalah, D., Khan, S.A., Al-Ashram, Y. et al., 2022. An integrated framework for the assessment of environmental sustainability in wood supply chains. In: Environmental Technology and Innovation, vol. 27, ID article 102429. DOI: 10.1016/j.eti.2022.102429.
7. Egghe, L., 1986. The dual of Bradford's law. In: Journal of the American Society for Information Science, vol. 37(4), pp. 246-255. DOI: 10.1002/(SICI)1097-4571(198607)37:4<246::AID-ASI10>3.0.CO;2-D.
8. Gordeev, R., 2020. Comparative advantages of Russian forest products on the global market. In: Forest Policy and Economics, vol. 119, ID article 102286. DOI: 10.1016/j.forpol.2020.102286.
9. Kostrzewski, M., Chamier-Gliszczyński, N., Królikowski, T., 2020. Selected reflections on formal modeling in Industry 4.0. In: Procedia Computer Science. vol. 176, pp. 3293-3300. DOI: 10.1016/j.procs.2020.09.118.
10. Lima, N.E., Guimarães, R.A., Almeida-Júnior, E.B. et al., 2020. Temporal trends, impact and partnership in floristic and phytosociology literature in the Brazilian Cerrado. In: Flora, vol. 273, ID article 151721. DOI: 10.1016/j.flora.2020.151721.
11. Naranan, S., 1970. Bradford's Law of Bibliography of Science: an Interpretation. In: Nature, vol. 227(5258), pp. 631-632. DOI: 10.1038/227631a0.
12. Prathiba, S., Kumar, P.S., Vo, D.-V.N., 2022. Recent advancements in microbial fuel cells: A review on its electron transfer mechanisms, microbial community, types of substrates and design for bio-electrochemical treatment. In: Chemosphere, vol. 286, ID article 131856. DOI: 10.1016/j.chemosphere.2021.131856.
13. Santos, A., 2020. RPubs - Psychomotricity – Bibliometrics. Available at: https://rpubs.com/andrelmfsantos/psychomotricity. Accessed on: May 20, 2023.
14. Viju, V.G.W., 2013. Application of

Bradford's law of scattering to the literature of library & information science: A study of doctoral theses citations submitted to the Universities of Maharashtra, India. In: Library Philosophy and Practice (e-journal), vol. 1054, Available at: https://digitalcommons.unl.edu/libphilprac/1054. Accessed on: 11 April, 2023.

15. Vivien, F.-D., Nieddu, M., Befort, N. et al. 2019. The Hijacking of the bioeconomy. In: Ecological Economics, vol. 159, pp. 189-197. DOI: 10.1016/j.ecolecon.2019.01.027.