Acta Biologica Plantarum Agriensis
ISSN 2061-6716 (Print), 2063-6725 (Online)
https://ojs.uni-eszterhazy.hu/index.php/ABPA
https://doi.org/10.21406/abpa.2023.11.1.129

11(1): 129–137 (2023)
Dedicated issue
Research article

A BIBLIOMETRIC ANALYSIS OF THE PUBLICATIONS BY TAMÁS PÓCS

János Podani

Department of Plant Systematics, Ecology and Theoretical Biology, Institute of Biology, Eötvös University, Pázmány P. s. 1/c, H-1117 Budapest, Hungary; and Institute of Evolution, Centre for Ecological Research, Budapest, Hungary; E-mail: janospodani@gmail.com

Abstract: This paper provides a statistical evaluation of the publication record of Tamás Pócs, from 1954 to 2022. During this period, he contributed to 550 publications of which 232 were authored only by himself. Frequency histograms are used to demonstrate the distribution of the number of his articles, books and book chapters for six variables: topic, document type, number of authors, year of appearance, geographical scope and language. It is remarkable that his publication activity was the highest in the past 20 years. Principal components analysis is used to reveal the correlation structure between ten features characterizing the publications.

Keywords: bryology, citations, d-correlation, nominal variables, principal component analysis, Scimago rank

INTRODUCTION

On the occasion of the 90th birthday of Tamás Pócs, to whom this volume of the journal is dedicated, nothing is more timely than a summary of his enormous scientific output. His first article appeared in 1954 in the Hungarian botanical journal, *Botanikai közlemények*, when he was still student at Eötvös Loránd University, Budapest. Since then, according to the database of the Hungarian Academy of Sciences (MTMT, "Storage for the Hungarian Scientific Publications") the number of his papers, books, book chapters and other types of publications has reached 550. Information in MTMT, supplemented with bibliographic data that can be generated from the records, gives me an opportunity to celebrate him by a special paper, a bibliometric analysis of his work. That is, rather than analyzing his epiphyllous liverwort

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

assemblage data from the tropics (e.g. Pócs *et al.* 2020) I change the focus and replace the bryophyte species with his publications as objects of the study.

Bibliometrics is an approach for evaluating large volumes of research publications. Standard descriptive statistics, network analysis and multivariate methods are used most often to exhaust information from the data which is not apparent at first sight. In my contribution, I present some descriptive charts while placing emphasis on the relationships among features or variables that characterize each individual item in the database, i.e., a record of a given publication. The method to be used here is a novel procedure (Podani *et al.* 2023) for calculating the correlation between variables that are expressed on different measurement scales. Although the method is originally suggested for processing ecological data, I hope that the present approach will serve as an example for future studies in the rapidly expanding field of bibliometrics.

MATERIAL AND METHODS

The core material of the analysis was extracted from the MTMT database in November, 2022. In this, the basic information on each item included the authors, the title, the year and type of publication, plus the number pages and the number of citations. Some duplicates were removed and a few errors were corrected according to the bibliographic account on Tamás's life by Orbán and Pénzes-Kónya (2013). Finally, as mentioned above, 550 records remained in the data set. In addition to the classical bibliographic variables, I used further traits that are extremely important in characterizing the work of a scientist. These are the subject matter of the publication, the geographical scope, the language, and the country in which the paper appeared. I did my best to classify the publications; and in doubtful cases I have chosen the most appropriate category. I also considered whether the journal has Scimago rank, and if so I used the categorization determined for the year of 2022 (https://www.scimagojr.com). I decided to do so, because for the actual years of publication it was not always possible to find any rank score. However, I am convinced that the most recent ranking still provides a good approximation to the relative importance of the journals to which

Tamás has contributed. For many types of publication, such as book chapters and very local journals, no Scimago ranks could be determined at all. Thus, this trait is a representative of the so-called combined scale type which has a neutral state (unranked) plus four ranks, from 1 to 4 (1 standing for the most prestigious journals). *Table 1* gives a summary of the ten variables used, their scale type and also lists the possible states wherever appropriate.

Table 1. Bibliographic features used in the present study.

No.	Feature	Scale type	Possible values
1	**Publication year**	interval	integer, 1954–2022
2	**Number of authors**	ratio	integer, 1–
3	**Number of pages**	ratio	integer, 1–
4	**Number of citations**	ratio	integer, 0–
5	**Topic**	nominal	1–systematics; 2–bryoecology, biogeography; 3–zoology; 4–community ecology; 5–astrobiology; 6–floristics; 7–personalia; 8–book review; 9–educational
6	**Geographic scope**	nominal	1–general, global; 2–Europe; 3–Africa; 4–Asia; 5–Australia; 6–Americas
7	**Locality of publisher**	nominal, binary	1–foreign; 2–Hungary
8	**Language**	nominal	1–English; 2–Hungarian; 3–French; 4–German; 5–Polish; 6–Spanish; 7–Estonian
9	**Publication type**	nominal	1–journal paper; 2–book; 3–conference publication; 4–book chapter; 5–dissertation; 6–other conference material; 7–miscellaneous
10	**Scimago rank**	combined neutral/ordinal	0–unranked; 1 to 4–rank = Q1 to Q4

As seen in *Table 1*, the bibliographic variables have four different scale types. Their relationships were quantified using the d-correlation function (Podani *et al.* 2023) which is based on the comparison of difference matrices between objects (in this case, 550 objects, i.e. bibliographic items). Since the differences are easy to define for any scale type, the use of d-correlation circumvents the problem that available correlation functions do not apply to any combination of scale, especially the combined one.

Futher advantage of the method is that missing scores are also allowed (in this case only 5 data values were missing). The 10 × 10 correlation matrix was then subjected to principal components analysis, using the SYN-TAX 2000 package (Podani 2001), to derive an ordination of variables.

RESULTS AND DISCUSSION

Descriptive statistics

Summary statistics characterizing the research work conducted by Tamás over the years from 1954 to 2022 are presented in form of frequency histograms in *Figure 1*. It comes at no surprise that most of his publications are devoted to systematics, bryoecology and biogeography (a total of more than 350). However, his wide interest in other disciplines is demonstrated by more than 50 publications in community ecology and, which is somewhat less expected, in astrobiology as well. Floristics, personalia and educational material are also represented by dozens of publications, and Tamás also had some (5) contributions to zoology. The distribution of his works over geographical regions is somewhat more even, with the maximum in the global/general category, followed by the publications devoted to the flora of his favourite continent, Africa. The other continents are also represented by significant numbers of papers. The distribution of the number of publications by language and by type is extremely skewed, that is, Tamás has followed the general practice in science and published his achievements in the English language in periodicals. However, his mothertongue, Hungarian was not neglected either! Also skewed is the histogram showing the number of authors. Although Tamás has nearly 250 publications of his own, demonstrating his excellence and independence in the field, he also loved to cooperate with 2–10 co-authors. The number of publications with more than 10 co-authors is much fewer, with a remarkable maximum of 68 in a joint report published in the *Journal of Bryology* (Ellis *et al.* 2021).

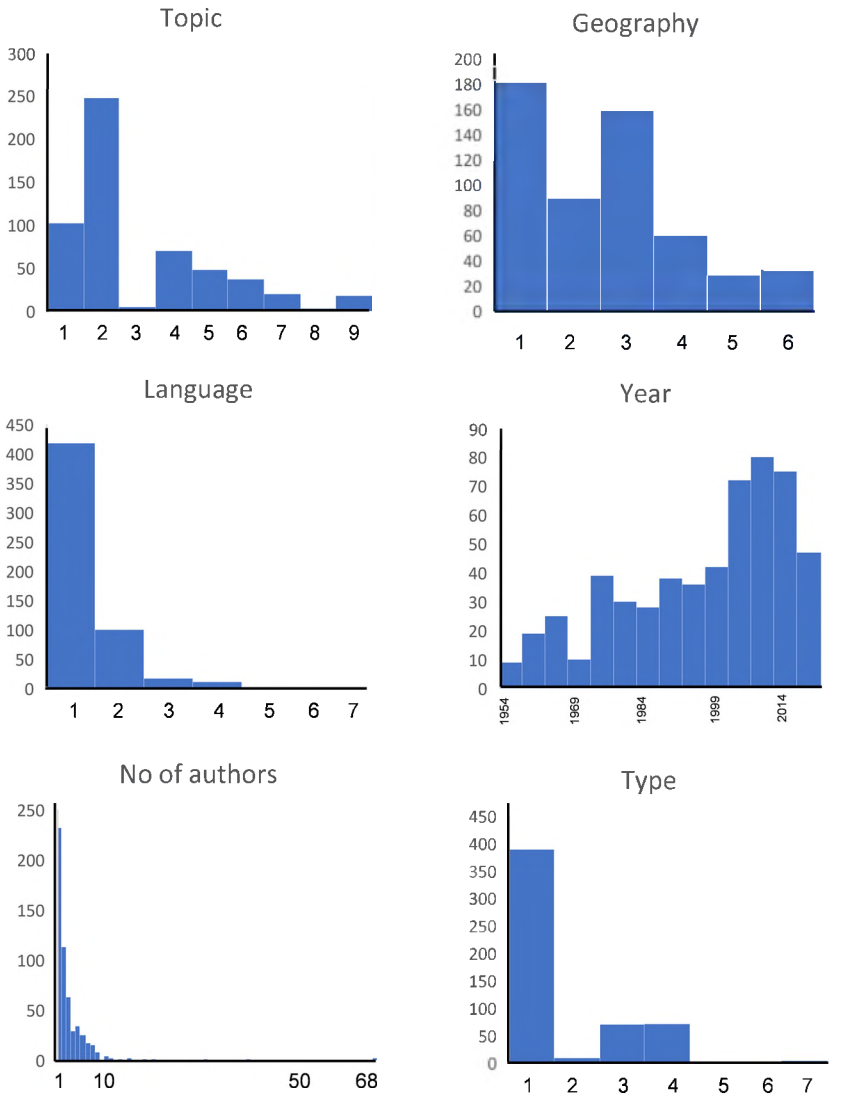

Figure 1. Histograms showing the distribution of the number of publications by Tamás according to six criteria. See *Table 1*, for feature codes shown on the horizontal axes. Data for publication year were accumulated into four-year wide categories.

The number of publications shown in the function of year is perhaps the most interesting of all diagrams. In the early periods of his life as a naturalist, Tamás had an average of 5 articles per year,

which increased to 10 and remained there until the early 2000's. Then, we can observe a jump up to an annual mean of 20 (i.e. 80 per 4 years, *Figure 1*). In other words, his research output greatly increased after his retirement, when he was allowed to allocate more time to research work than in his earlier years. And we can be happy to realize that his activity did not change much near the age of 90!

Correlation of features

The second most interesting issue in the bibliometric evaluation of the publications of scientists is the comparison of features. The matrix of d-correlations between the ten variables includes many values ranging between –0.1 to 0.1, larger correlations are less common. Although sample size is very high and therefore even these low values can be significant, I shall restrict interpretation to higher values. The strongest correlation (r = 0.64) is between the number of citations and the number of pages, and these two are also highly correlated with the number of authors (r = 0.42 and r = 0.28, respectively). These three ratio-scale features appear as a separate group along the first three PCA dimensions (*Figures 2a–b*). It is interesting, however, that such extremely high correlations are due to a single – albeit most important – publication written by 41 authors (Söderström *et al.* 2016) which takes 828 pages in *Phytokeys*, with 302 citations (in November, MTMT). The latter two numbers are the highest in the publication record of Tamás.[1] If we disregard this single publication and redo the analysis, the above correlations are lowered to 0.1, –0.02 and 0.23, respectively. In the PCA ordination, therefore, the number of citations and the number of authors remain relatively close, whereas the number of pages is shifted away (*Figures 2c–d*). This phenomenon shows the sensitivity of the analysis to a variable with an extreme value, i.e. to an outlier, because the second longest publication (Lovett and Pócs 1993) had "only" 300 pages. Therefore, even standardization implied in the d-correlation was not influential enough. Also, the question arises whether the number of pages is commensurable,

[1]Note that on January 10, 2023, the total number of citations reached 350 in MTMT, reflecting a rapid increase in this statistic.

because the actual contents and the size of a page may vary considerably over journals and books.

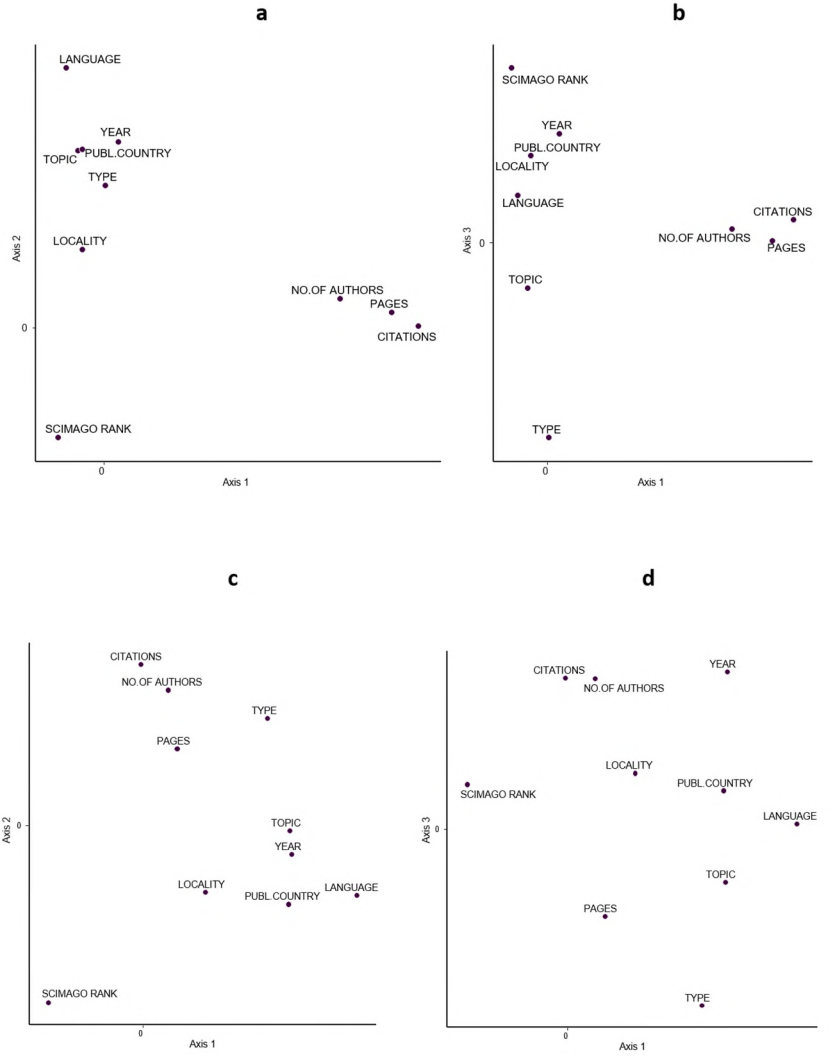

Figure 2. Pincipal components analysis of bibliometric variables by considering all publications (**a**–**b**) and by removing the most often cited item (**c**–**d**). Left: axes 1 and 2, right: axes 1 and 3. Eigenvalues, which reflect the variances accounted for by each component, are 19%, 17% and 13% for a–b, and 17%, 14% and 13% for c–d.

The language and the country of publication are highly correlated (r = 0.31), an expected result due to the concentration of Hungarian articles. Language and year have the same correlation, reflecting the increase of English language papers over the years. The Scimago rank has negative correlations with some variables, and is thus positioned relatively far from the others in the ordination. It has r = –0.18 with the number of citations (reduced data set) which reflects the trend that lower ranks (low Q values, i.e., "better" journals) are associated with more citations. The highest negative correlation, r = –0.31 is between Scimago rank and the type of the publication. This is due to two factors: large differences in ranks can only appear within the same, i.e. the journal paper category (389 items), whereas the other six types of publications agree that they are all unranked. These relationships do not necessarily show up perfectly in the ordinations, because the first three dimensions explained less than 50% of the total variance.

Acknowledgement – I am honoured to be present as co-author in a few publications by Tamás, thus contributing a little to his success. I am grateful to Judit Végh (András Pető Faculty, Semmelweis University, Budapest) for her assistance in extracting the bibliographic data from MTMT.

REFERENCES

Ellis, L.T., Ah-Peng, C., Aslan, G., Bakalin, V.A., Bergamini. A., Callaghan, D.A., Campisi, P., Raimondo, F.M., Choi, S.S., Csiky, J., Csikyné Radnai, E., Cykowska-Marzencka, B., Czernyadjeva, I.V., Kalinina, Y.M., Afonina, O.M., Domina, G., Drapela, P., Fedosov, V.E., Fuertes, E., Gabriel, R., Kubová, M., Soares Albergaria, I., Gospodinov, G., Natcheva, R,. Graulich, A., Hedderson, T., Hernández-Rodríguez, E., Hugonnot, V., Hyun, C.W., Kirmaci, M., Çatak, U., Kubešová, S., Kučera, J., La Farge, C., Larraín, J., Martin, P., Mufeed, B., Manju, C.N., Rajesh, K.P., Németh, C., Nagy, J., Norhazrina, N., Syazwana, N., O'Leary, S.V., Park, S.J., Peña-Retes, A.P., Rimac, A., Alegro, A., Šegota, V., Koletić, N., Vuković, N., Rosadziński, S., Rosselló, J.A., Sabovljević, M.S., Sabovljević, A.D., Schäfer-Verwimp, A., Sérgio, C., Shkurko, A.V., Shyriaieva, D., Virchenko, V.M., Smoczyk, M., Spitale, D., Srivastava, P., Omar, I., Asthana, A.K., Staniaszek-Kik, M., Cienkowska, A., Ştefănuţ, M.M., Ştefănuţ, S., Tamas, G., Bîrsan, C.C., Nicoară, G.R., Ion, M.C., Pócs, T., Kunev, G., Troeva, E.I., van Rooy, J., Wietrzyk-Pełka, P., Węgrzyn, M.H., Wolski, G.J., Bożyk, D. & Cienkowska, A. (2021). New national and regional bryophyte records, 65. *Journal of Bryology* **43**(1): 67–91. https://doi.org/10.1080/03736687.2021.1878804

LOVETT, J.C. & PÓCS, T. (1993). *Assessment of the condition of the Catchment Forest Reserves, a botanical appraisal.* Ministry of Tourism, Natural Resources and Environment, Dar es Salaam.

ORBÁN S. & PÉNZES-KÓNYA E. (2013): Professor Tamas Pócs. The record of his 80 years. *Polish Botanical Journal* **58**(1): 3–30. https://doi.org/10.2478/pbj-2013-0001

PÓCS, T., LEE, G.E., PODANI, J., PESIU, E., HAVASI, J., TANG, H.Y., MUSTAPENG, A.M.A. & SULEIMAN, M. (2020). A study of community structure and beta diversity of epiphyllous liverwort assemblages in Sabah, Malaysian Borneo. *PhytoKeys* **153**: 63–83. https://doi.org/10.3897/phytokeys.153.53637

PODANI, J. (2001). SYN-TAX 2000. *Computer programs for data analysis in ecology and systematics*. User's Manual. Scientia, Budapest.

PODANI, J., SCHMERA, D. & BAGELLA, S. (2023). Correlating variables with different scale types: a new framework based on matrix comparisons. *Methods in Ecology and Evolution* **14**(4): 1049–1060. https://doi.org/10.1111/2041-210X.14074

SÖDERSTRÖM, L., HAGBORG, A., VON KONRAT, M., BARTHOLOMEW-BEGAN, S., BELL, D., BRISCOE, L., BROWN, E., CARGILL, D.C., COSTA, D.P., CRANDALL STOTLER, B.J., COOPER, E.D., DAUPHIN, G., ENGEL, J.J., FELDBERG, K., GLENNY, D., GRADSTEIN, S.R., HE, X., HEINRICHS J., HENTSCHEL, J., ILKIU-BORGES A.L., KATAGIRI, T., KONSTANTINOVA, N.A., LARRAÍN, J., LONG, D.G., NEBEL, M., PÓCS T., PUCHE, F., REINER-DREHWALD, E., RENNER, M.A.M., SASS-GYARMATI, A., SCHÄFER-VERWIMP, A., MORAGUES J.G.S., STOTLER, R.E., SUKKHARAK, P., THIERS B.M., URIBE, J., VÁŇA, J., VILLARREAL, J.C., WIGGINTON, M., ZHANG, L. & ZHU, R.L. (2016). World checklist of hornworts and liverworts. *Phytokeys* **59**: 1–828. https://doi.org/10.3897/phytokeys.59.6261

(submitted: 25.01.2023, accepted: 05.03.2023)