

# A Normalized Index to Quantify Publication Productivity across Disciplinary Boundaries\*

Koh Yamamoto<sup>†</sup>, Masumi Ishikawa<sup>†</sup>

## Abstract

The number of papers and number of citations have been widely used as indices for research capability in various situations, such as the comparison of research institutions and performance evaluation of researchers. However, it is well known that these naïve indices have large disciplinary variations, making them difficult to apply across disciplinary boundaries. Although various normalized citation indices have been provided by research assessment tools, no widely accepted indices have been established for the number of papers despite this number's significance in research performance metrics. In the present paper, we propose a normalized index for the number of papers to enable a fair comparison of research capability. A key idea is to introduce the concept of work efficiency into the quantification of publication productivity. We further investigate the effects of normalization on the publication data of our institute. The normalization resulted in a relative change in the total number of papers in inverse relation to publication intensity in each discipline. Similar results were obtained between two major bibliometric databases for the publication productivity in well-populated disciplines with similar field coverage. Bootstrap analysis revealed that a sample size of 200–300 is required to obtain statistically significant publication productivity.

*Keywords:* Disciplinary variation, Normalization of the number of papers, Publication productivity, Publication efforts

## 1 Introduction

In recent years, research institutions have made every effort to enhance the capabilities of education and research in a highly competitive environment; the research area is the focus of this paper. In order to enhance research capability, institutions urgently need to evaluate their performance correctly at multiple levels, i.e., the level of the whole institution, department, and individual researcher. The number of papers and number of citations are generally used as indices of publication productivity and impact, respectively, and the combination of the two indices represents research performance. As suggested by the numbers in Table 1, however, there is a large disciplinary variation in the bibliometric indices [1][2].

---

\* An earlier version of this paper was presented at the 5th Meeting on Japanese Institutional Research, Kumamoto (12–13 July 2016).

<sup>†</sup> Kyushu Institute of Technology, Fukuoka, Japan

Table 1: Characteristics of bibliometric data over diverse disciplines. The number of papers ( $N_p$ ), number of authors per paper ( $N_a/N_p$ , where  $N_a$  is the number of authors involved), and citation rate (CR) are derived from the Essential Science Indicators of the Web of Science. Data were selected to represent both articles and reviews, and publication years 2010–2014 (March 10, 2016, date last accessed).

Disciplines	$N_p$	$N_a/N_p$	CR
Clinical Medicine	1,262,028	2.87	8.10
Chemistry	759,247	2.11	9.83
Engineering	554,365	2.02	5.33
Physics	554,001	2.17	8.04
Social Sciences, general	421,163	1.89	4.04
Materials Science	344,724	2.43	8.08
Biology & Biochemistry	342,015	3.68	10.55
Plant & Animal Science	339,737	2.62	5.66
Neuroscience & Behavior	241,327	3.24	11.25
Molecular Biology & Genetics	210,943	4.58	15.59
Environment/Ecology	202,152	2.87	8.06
Geosciences	201,970	2.35	7.56
Mathematics	199,798	1.11	2.67
Agricultural Sciences	192,282	2.85	5.32
Psychiatry/Psychology	184,443	2.17	7.26
Pharmacology & Toxicology	181,708	3.69	8.21
Computer Science	161,317	2.17	4.90
Economics & Business	125,865	1.39	4.52
Immunology	120,654	4.54	12.16
Microbiology	97,229	3.84	9.76
Space Science	68,714	2.37	12.81
Multidisciplinary	9,691	5.66	9.41

Research institutions usually differ in their coverage and emphasis of a discipline area, and the variation makes it difficult to draw an unbiased and reliable comparison of their research performance. The same is true of comparisons at the level of the department and individual researcher as long as there is a difference in the disciplines of the target for comparison. Therefore, there is a strong demand for the establishment of alternative metrics that take into account the disciplinary variation.

The number of citations is considered to be the manifestation of paper visibility, which, in turn, is an approximate index of paper quality. However, the number of citations tends to differ, not only among disciplines but also among publication years; the earlier papers are published, the more citations they have [3]. In consideration of this problem, research assessment tools, such as InCites of Thomson Reuters and SciVal of Elsevier, have provided normalized citation indices to users, e.g., Category Normalized Citation Impact (CNCI) of Thomson Reuters [4] and Field Weighted Citation Index (FWCI) of Elsevier [5]. These in-

dices are calculated by normalizing the number of citations so that the average becomes one over a given time period in a specific group of papers with the same discipline, publication year, and document type. This enables us to make a more reliable comparison of research impact.

In contrast, the development of indices for the number of papers has not been addressed actively, as mentioned by [6], probably because the index measures aspects of quantity regardless of aspects of quality. It is well known that in the 1990s, Australia experienced a considerable increase in its share of the global publication output while simultaneously facing a drop in its share of the global citation pool; [7] ascribed this phenomenon to quantitative measures employed in allocating core research funds in Australian universities. Nevertheless, the number of papers has important advantages over the citation index; the latter is prone to a lack of information for the publications not indexed in major citation databases, and reflects research conducted in the past often a considerable number of years prior to the time of evaluation [8]. For this reason, the number of papers has still been included in metrics on research performance evaluation [9]. In Japan, the number of papers has been used for the performance evaluation of individual researchers in over 80% of national universities [10], suggesting that the index is taking hold at a practical level. The index is also used for domestic or international benchmarking of research performance (e.g., [11][12]) and the performance-based research funding systems in combination with a series of agreed indicators or extra weight given to the publications in most selective international journals, series and book publishers [13] [14]. We also maintain the viewpoint that the number of papers is no more than a part of metrics on research performance evaluation; practical evaluation should be conducted from diversified perspectives such as publication activity and impact, patents, conference presentations, and databases.

Numerous studies have compared researchers within the same discipline to avoid distortions in productivity rankings caused by disciplinary variation in publication productivity. A widely known index called *h*-index, which estimates research performance using a combination of the number of papers and number of citations [15], is also disciplinary-dependent [1]; developing more sophisticated indices for the number of papers is crucial to draw a comparison across disciplines. To the author's knowledge, only the study by [16] has proposed the index for the number of papers normalized with publication productivity within each discipline before aggregating the data. In general, it is considered that the number of papers per researcher in a specific discipline represents the publication productivity of the discipline. However, the number of papers that a single researcher can publish in a given time period is restricted by the working hours he or she can afford to spend. If a researcher devotes most of his or her publication efforts to a specific discipline, he or she can publish only a limited number of papers in other disciplines; the publication effort is here defined as an individual's effort devoted to a research field following [17]. Therefore, the number of papers per researcher represents not only the publication productivity but also the extent to which involving researchers devoted publication efforts to the discipline. The novelty of the present study resides in the introduction of the concept of work efficiency into the quantification of publication productivity, and it is not applied to the calculation of normalized citation indices such as CNCI and FWCI.

In the present study, a normalization method for the number of papers will be proposed with the aim of fairly comparing publication productivity across disciplinary boundaries. To enable an evaluation of the adequacy and usefulness of the normalized index, publication data should be prepared carefully to ensure correct affiliation and aggregation. Currently, it is difficult to collect reliable data on researchers all over the world, or even those in Japan.

Furthermore, because the total population of researchers in each discipline is not known, it is difficult to estimate the disciplinary average of publication productivity, which is crucial for the method. Therefore, the method will be applied to the publication data of the authors' institution of affiliation, Kyushu Institute of Technology (hereinafter referred to as Kyutech), for which a time-consuming manual verification process for the data is possible and the total population is clearly known. It is unreasonable to assume the disciplinary average for the researchers of one institution as a proxy of that for the total population. This paper is therefore limited to presenting the methodology (Section 2) and soundness of the proposed normalized index in light of variations in the disciplinary averages of the publication productivity of Kyutech (Section 3.1), as well as investigating the minimal sample size required for accurate determination of the productivity (Section 3.2). Accuracy aside, effects on a researcher's performance evaluation is also shown as an example of application of the normalized index (Section 3.3). Finally, we discuss the issues and challenges for the future revealed in the course of this study (Section 3.4).

## 2 Materials and Methods

### 2.1 Data

Table 2 shows a brief description of the publication data used in the present study. The primary data sources are the Science Citation Index Expanded of the Web of Science (Thomson Reuters) and Scopus (Elsevier), and document types were limited to article and review. In addition, data were extracted from these publication databases not for a single year, but for five years (2010–2014) in consideration of the delay in registration of papers to the databases and for the purpose of making the result more robust by increasing the amount of data. For the Web of Science, the number of papers that met the conditions above and the number of authors affiliated with Kyutech were 1430 and 288, respectively. These values were greater for Scopus (1653 and 354) reflecting that journal coverage is wider in Scopus than in the Web of Science [18]. To ensure the data quality, the authors were restricted to tenure faculty members; non-tenure faculty members and students were excluded from the analysis. Name disambiguation was conducted for the Web of Science by requesting

Table 2: A brief description of the Kyutech's publication data.

	Thomson Reuters	Elsevier
Database	Web of Science, SCIE*	Scopus
Disciplines	22 of ESI†	27 of ASJC‡
Document types	Article, Review	Article, Review
Published year	2010–2014 (5 years)	2010–2014 (5 years)
Number of papers	1430	1652
Number of authors	288	354

\* Science Citation Index Expanded

† Essential Science Indicators

‡ All Science Journal Classification

confirmation of the publication list from all authors, while only automatic disambiguation is made by the provider for Scopus.

## 2.2 Definition and assumptions

To take the work efficiency (see Section 1) into consideration, a unit of publication effort is defined as manpower, with each researcher having a total of one manpower in a given time period (2010–2014 in the present study). Furthermore, to apply our idea to bibliometric data, we adopt the following simplifying assumptions: (1) coauthors equally contribute to the publication, (2) the paper count can be allocated evenly if the paper is assigned to multiple disciplines, (3) a given researcher's publication efforts devoted to a specific discipline are proportional to the ratio of the number of papers in the discipline to the total number of his or her papers, and (4) non-publishing researchers, whose fields of study are not identifiable, can be ignored.

Based on the first assumption, the paper count of a joint paper was equally partitioned and allocated to the authors. We here applied a fractional counting method [19]. To evaluate the degree of contribution to the publication performance of our university, the number of authors to be used in the partitioning was defined, not as all the authors, but as authors belonging to Kyutech. For example, if two authors from Kyutech and one author from another research institution collaborated in the joint paper, the fractional count of 0.5 was allocated to each Kyutech author. Furthermore, based on the second assumption, partitioning was done for disciplines in the case of the Scopus data; many journals are assigned to more than one discipline in ASJC, unlike in ESI of the Web of Science (see Table 2). If the journal in which the joint paper was published is assigned to two disciplines, the value of 0.5 mentioned above was further divided by two, and each Kyutech author was given the fractional count of 0.25 in both the disciplines. We had no choice but to simplify the allocation method because it is difficult to know the real weight of individual authors' publication efforts devoted to the publication and proximity of the field of study to the assigned disciplines.

## 2.3 Normalization

The method to calculate the normalized number of papers, a new index named Discipline Weighted Publication Productivity (DWPP), is described below. First, the publication effort ratio ( $\beta$ ) is calculated for individuals as in (1):

$$\beta_{ij} = \frac{\alpha_{ij}}{\alpha_i} \quad (1)$$

where  $\alpha$  is the number of papers and subscripts  $i$  and  $j$  indicate the identity numbers of the author and discipline, respectively. The symbol  $\alpha_i$  represents the total number of papers for author  $i$ , being integrated over all disciplines. Next, publication productivity of discipline  $j$  ( $P_j$ ) is calculated as shown in (2):

$$P_j = \frac{\sum_{i=1}^N \alpha_{ij}}{\sum_{i=1}^N \beta_{ij}} \quad (2)$$

where  $N$  is the total number of authors included in the analysis. The numerator and denominator represent the total number of papers and the total effort ratio, respectively, in discipline  $j$ . The DWPP of author  $i$  ( $\gamma_i$ ) is calculated by dividing the number of papers by the reference value  $P_j$  in each discipline and integrating the ratios over all disciplines, as in (3):

$$\gamma_i = \sum_{j=1}^M \frac{\alpha_{ij}}{P_j} \quad (3)$$

where  $M$  is the total number of disciplines in the database: 22 for the Web of Science and 27 for Scopus (see Table 2). The  $\gamma_i$  values of all authors average one; if a given author has  $\gamma_i$  larger than one, he/she has above-average publication productivity.

## 2.4 Confidence interval

Determination of  $P_j$  is a core of the normalization, and its accuracy needs to be ensured. In the present study, we calculated confidence intervals of the random variable  $P_j$  by the computer-intensive method of statistical analysis known as the “bootstrap method.” This method is a kind of Monte Carlo simulation; the property of the parent population is estimated with the use of numerous bootstrap samples, which are made up of a single sample by resampling with replacement. Because the method needs not presume a probability density function of the estimator or a parametric model [20], it is applicable to the present study, in which available data are not ample and the distribution of the estimator  $P_j$  is not known.

Using a dataset consisting of the number of papers and publication effort ratios for individuals, which are counted in each discipline, the confidence interval of  $P_j$  is calculated through the following procedure:

- (1) Create a sample of arbitrary size  $n$  by random sampling from the original dataset.
- (2) Construct 2000 resamples with replacement using a sample size of  $n$ .
- (3) Calculate  $P_j$  values from the bootstrap resamples to obtain a bootstrap distribution.
- (4) Obtain a 95% bootstrap percentile confidence interval.

The statistical procedures were implemented in R-software [21].

## 3 Results and Discussion

### 3.1 Disciplinary variation in $P_j$

Figure 1 shows the calculated values of  $P_j$ , which are sorted by the total number of authors in each discipline. Error bars indicate the 95% confidence interval obtained by the bootstrap method. The sample size  $n$  in the procedure (1) in Section 2.4 is the total number of authors here. Considering statistical errors, disciplines with fewer than 10 authors are excluded from consideration here, for the data of the Web of Science. The number of remaining disciplines is 11; the top 11 well-populated disciplines are also shown for the Scopus data. The minimal sample size required for the accurate determination of  $P_j$  will be described in Section 3.2.

The  $P_j$  values estimated from the Web of Science data averaged 4.79, and the standard deviation  $\sigma$  was 1.74. In light of the 95% confidence intervals, the  $P_j$  values were significantly higher than the average in disciplines such as Chemistry, Physics, and Biology

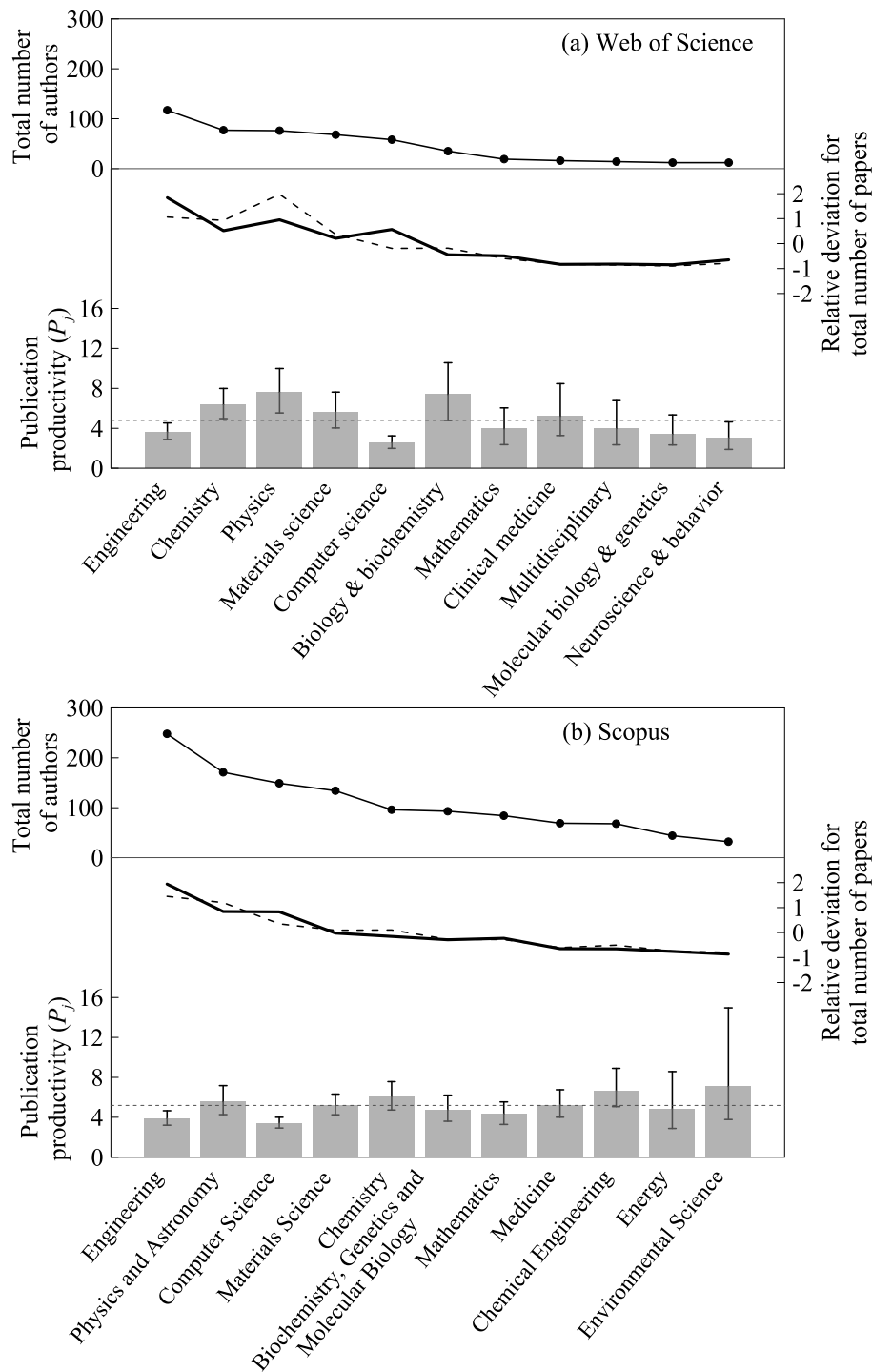


Figure 1: Disciplinary variation of the total number of authors (closed circles with solid line), publication productivity ( $P_j$ , bars), and relative deviation for actual (dashed line) and normalized (thick solid line) total number of papers. The relative deviation was calculated by dividing each value minus the mean by the mean. Thin dotted lines and error bars on the  $P_j$  bars indicate the average values of  $P_j$  over all disciplines in this figure and 95% confidence intervals calculated by the bootstrap method, respectively. The analysis was done for the data of (a) the Web of Science and (b) Scopus.

& Biochemistry (hereinafter referred to as high- $P_j$  disciplines). In contrast, significantly lower values were observed in Engineering, Computer Science, Neuroscience & Behavior (hereinafter referred to as low- $P_j$  disciplines). For the Scopus data, the average and  $\sigma$  of  $P_j$  were 5.19 and 1.12, respectively, and the high- $P_j$  discipline was absent, while the low- $P_j$  disciplines were Engineering and Computer Science in common with the Web of Science. Regardless of which database we used, the relative deviation of the total number of papers was smaller for the normalized values than for the actual values in the high- $P_j$  disciplines, and the opposite was true in the low- $P_j$  disciplines. This suggests that the normalization worked as expected, according to the intention to make a relative increase in the number of papers in disciplines with low publication productivity.

To further investigate the similarity of  $P_j$  values between the two databases, we compared the  $P_j$  values of the top five well-populated disciplines in the Web of Science to those of the corresponding disciplines in Scopus (Figure 2). Despite slight differences in the magnitude, the observed variation in the  $P_j$  values was fairly similar, except for the pair of Physics in the Web of Science and Physics and Astronomy in Scopus. Analysis of the Kyutech's publication data shows that the results do not differ significantly between the two databases as far as the well-populated disciplines with similar field coverage are concerned. However, the robustness of the conclusion, should be further studied because our analysis was made mainly for disciplines included in a broad field, named Natural Sciences and Engineering; it has a relatively small difference in field coverage between the two databases in comparison to the other fields, i.e., Biomedical Research, Social Sciences and Arts and Humanities [18].

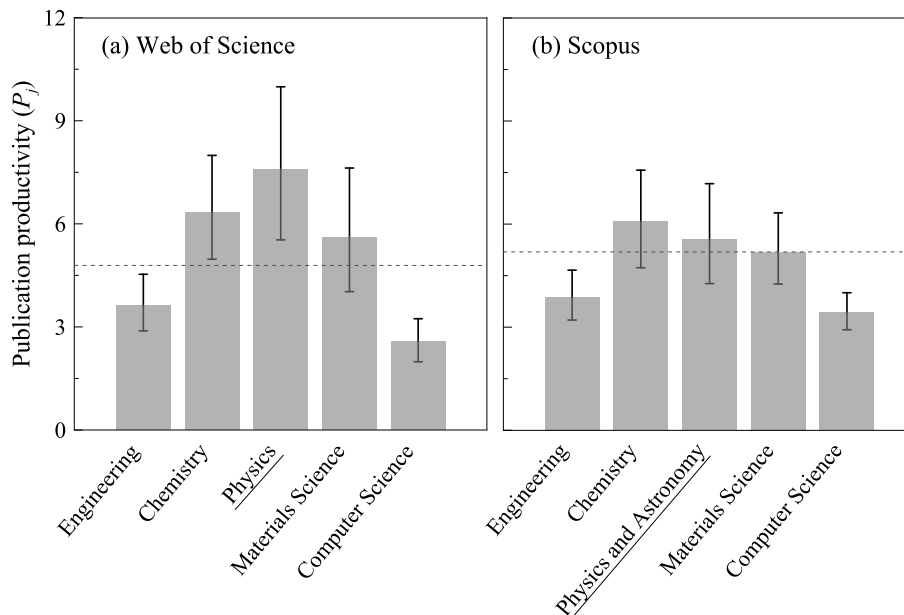


Figure 2: Publication productivity  $P_j$  for (a) the Web of Science and (b) Scopus. The top five well-populated disciplines of the Web of Science and the corresponding disciplines of Scopus are shown here. Underlined axis labels indicate the disciplines, field coverage of which is not in complete correspondence between the two databases. For thin dotted lines and error bars on the  $P_j$  bars, see the caption to Figure 1.



It is worth pointing out here that every researcher is equally assumed to have one manpower in the analysis (see Section 2.2). In fact, the number of hours that a given researcher can devote to a research field depends on his or her commitment to other tasks, e.g., activities for education and organizational operation. It would not be appropriate to compare the  $P_j$  values if the average conditions vary to a large extent among target groups.

### 3.2 Minimal sample size for $P_j$ determination

It is crucial to determine the confidence interval of  $P_j$  with high accuracy for a comparison of the values among disciplines. As long as the minimal sample size required for accurate determination is unknown, it is difficult to guarantee that  $P_j$  values in Figure 1 are statistically significant. To estimate the minimal sample size, we take all the publication data as a parent population; the sample size is 542 for the Web of Science and 1340 for Scopus. The sample size substantially exceeds the number of authors (see Table 2) because many researchers are publishing papers in multiple disciplines. The difference is larger for the Scopus data because of an apparent increase in the amount of data caused by the allocation of paper count to multiple disciplines (see Section 2.2).

Figure 3 shows the 95% confidence intervals calculated by incrementing  $n$  in the bootstrap calculation procedure (1) (see Section 2.4) by 20; the  $n$  value starts from 20 to the extent that  $n$  does not exceed the sample size of the parent population. Because the results vary depending on which data are selected in the procedure (1), the calculation was made 100 times at each sample-size step with changing random seeds to obtain the average and  $\sigma$  of the upper and lower limits. As expected, a small sample size was accompanied by a large confidence interval, and the interval became narrower as  $n$  increased. The rate of change, defined as  $(x_k - x_{k-1})/x_{k-1} \times 100$  (%), where  $x$  indicates arbitrary variable and  $k$  iteration step, was calculated for the average of the limits to detect convergence to a steady state; it fell below 1% when the sample size exceeded 200 for the Web of Science data, while the threshold was 1.5 times larger, around 300, for the Scopus data. The difference may be attributable to the apparent increase in the amount of the Scopus data, mentioned above.

### 3.3 Effects on researcher rankings

Statistical analysis in the previous subsection proved that various magnitudes of uncertainty existed in the  $P_j$  values because of insufficient data. Furthermore, the resulting  $P_j$  is the disciplinary average for the researchers of Kyutech; there would be some deviations in the normalized index DWPP calculated with the  $P_j$  from its correct values calculated with the disciplinary average for the total population. Nevertheless, it is worth presenting a concrete example of application of the index to researchers' performance evaluation in order to suggest predictable consequences.

The change of the measure from the number of papers to the DWPP resulted in a change in researcher rankings based on publication productivity (Figure 4). The sparsely populated disciplines, which had been excluded in Section 3.1, were included in this analysis to provide all those concerned their rankings, disregarding possible large statistical errors. In both cases of using the data from the Web of Science and Scopus, the higher or the lower the researcher ranking before the normalization was, the smaller the change in ranking was. The normalization caused a large change in ranking for the middle-ranking researcher (in the top 21%–80% before normalization); the maximal variation width was  $-80$ – $60$  based on the Web of Science data, while it was  $-60$ – $50$  based on the Scopus data. The slightly

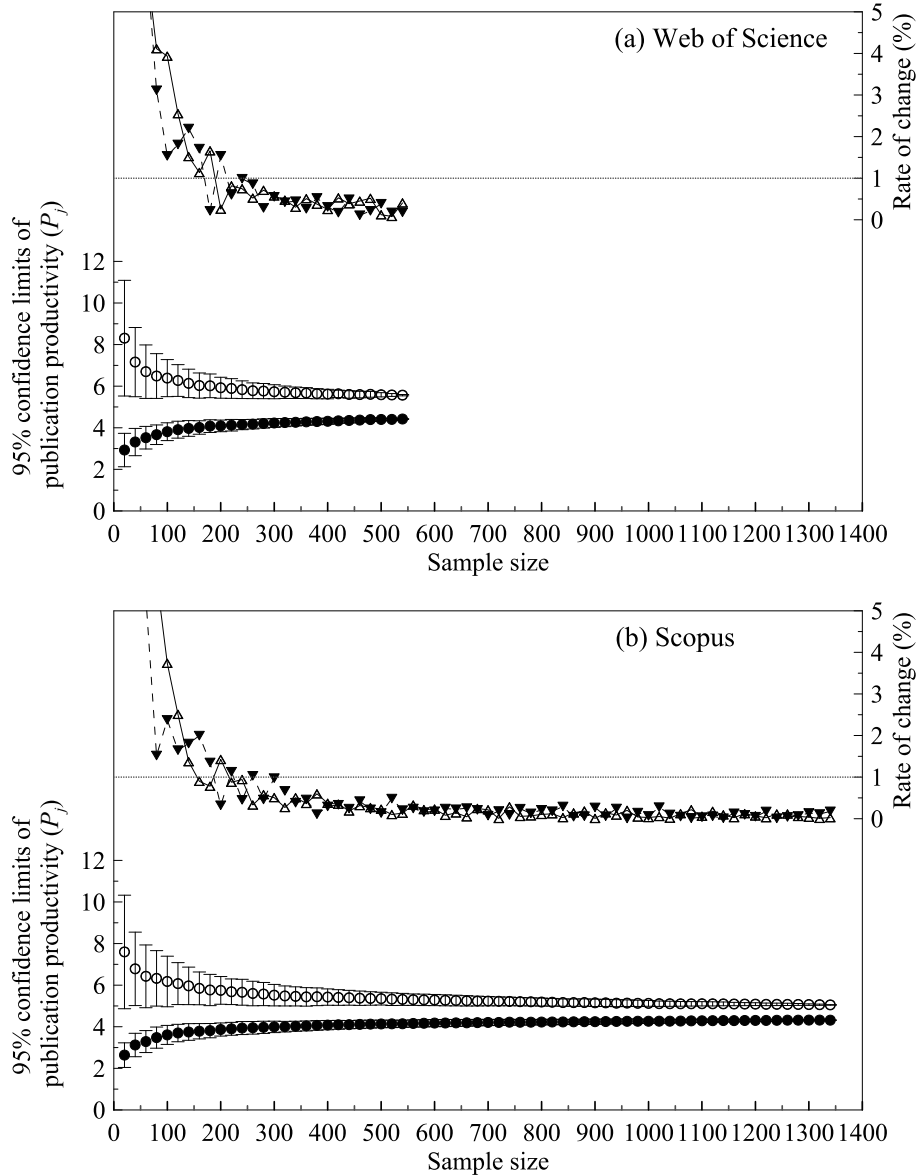


Figure 3: Dependence of 95% confidence limits of publication productivity  $P_j$  on the sample size. The average of upper limits (open circles) and lower limits (closed circles) were obtained from 100 times repetition of bootstrap calculations, and error bars indicate the standard deviation. Also plotted are the rates of change in the average of upper limits (open triangles) and lower limits (closed triangles). The values were calculated for the data of (a) the Web of Science and (b) Scopus.

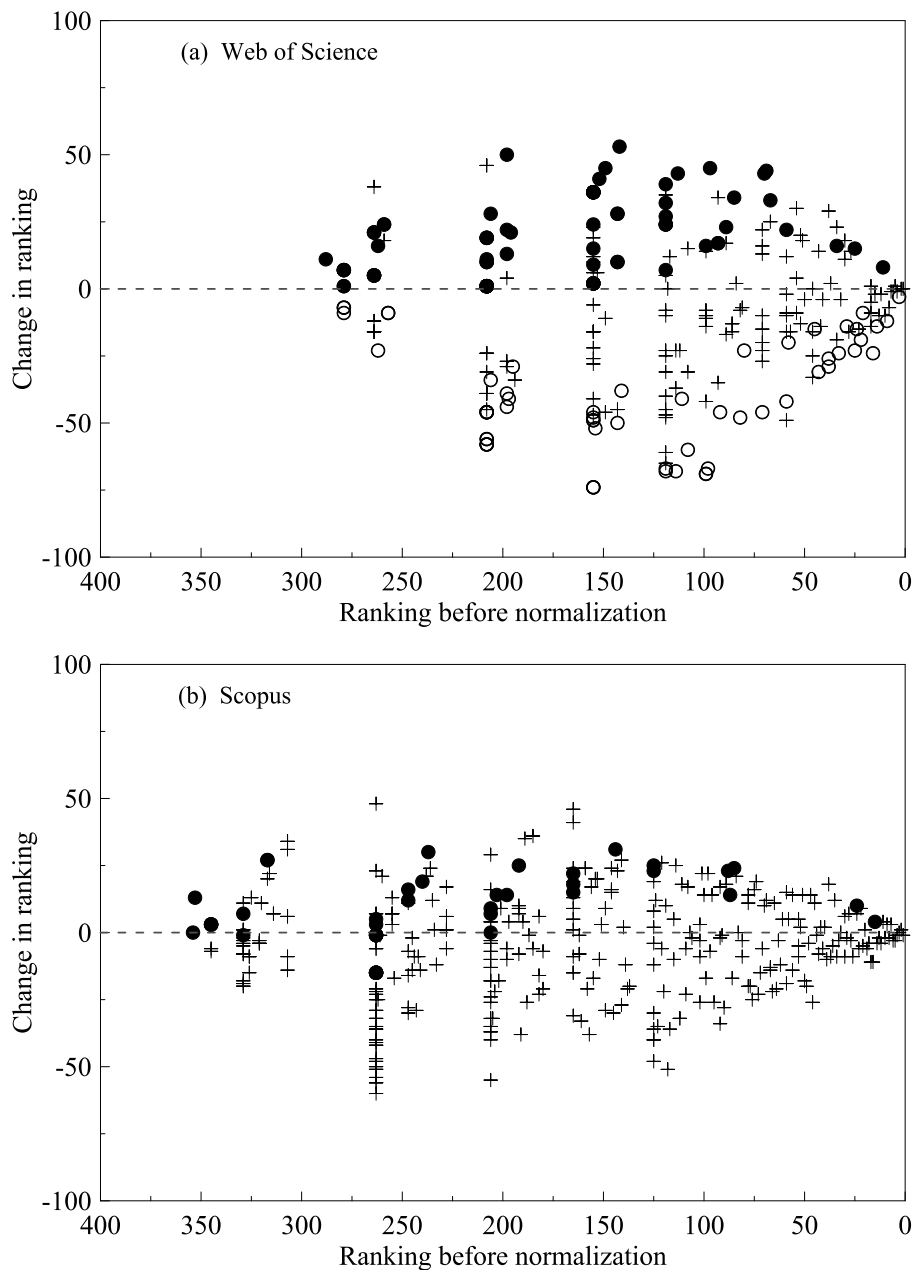


Figure 4: Change in researcher ranking calculated by subtracting the ranking before normalization from the ranking after normalization. Data are sorted by the ranking before normalization and plotted separately for researchers publishing papers mainly in the low- $P_j$  disciplines (closed circles), researchers publishing papers mainly in the high- $P_j$  disciplines (open circles), and otherwise (crosses). The analysis was done for the data of (a) the Web of Science and (b) Scopus.

small width for Scopus may be attributable to the larger number of disciplines of ASJC (see Table 2) or the larger number of disciplines per researcher caused by the allocation of paper count. In summary, the ranking of researchers publishing a moderate number of papers is changeable by normalization, and that of researchers who are prominently productive or unproductive is rather stable.

In every band of the ranking before normalization, most researchers who mainly published papers in the low- $P_j$  disciplines, with the publication effort greater than 90%, enjoyed a marked rise in ranking, and the opposite can be said for researchers with the great effort devoted to the high- $P_j$  disciplines, compared to the other researchers (Figure 4). Although they were few, some of the other researchers moved up in the ranking more than researchers publishing papers mainly in the low- $P_j$  disciplines (see the upper envelope of the plots in Figure 4a and b). By examining the details of the publication data, we found such researchers publishing in disciplines with  $P_j$  well below its average, but the  $P_j$  was insignificant in light of a large confidence interval; in some cases, the discipline was sparsely populated and had been excluded in Section 3.1 (e.g., Geoscience in the Web of Science, and Neuroscience and Arts and Humanities in Scopus). A small minority of the other researchers moved down in the ranking more than researchers publishing papers mainly in the high- $P_j$  disciplines (see the lower envelope of the plot in Figure 4a). They were publishing researchers in disciplines with statistically insignificant  $P_j$  well above its average (e.g., Microbiology in the Web of Science). These results suggest that, in determining individual rankings, it is crucial to determine  $P_j$  with accuracy in every discipline.

### 3.4 Challenges for the future

The most compelling work is to ensure a sufficient amount of publication data. In our analysis to determine the minimal sample size required for accurate determination of  $P_j$ , we found that the amount of data did not meet the requirements in every discipline. This has prevented us from making a firm conclusion about the disciplinary differences observed in  $P_j$  and using the normalized index DWPP for performance evaluation. We are now preparing to share information with other collaborative research institutes in Japan with the aim of increasing the amount of available data. Because the allocation of an individual's publication efforts to multiple disciplines is crucial for  $P_j$  determination, name disambiguation is an important task in the data collection. Furthermore, the data should include relevant information, such as position and affiliation, because recently researchers tend to work at more than one institution with different positions. There are a few more details to be studied: selecting a proper time period for the analysis and disciplinary classification, including document types other than article and review, and defining the authors to be included in the allocation of paper count.

There remains room for improvement in establishing a reliable index. First, we recognize that there is an intrinsic inaccuracy in the assumption on the allocation of an individual's publication efforts to multiple disciplines (see assumption (3) in Section 2.2); the proportional allocation is subject to an even difficulty of publication among disciplines at the individual level, which contradicts the disciplinary variation in  $P_j$  applied to normalize the number of papers. Although the inaccuracy was assumed to be ignorable in the results of this study, this is a future issue to be solved. Another important assumption we made regards the negligible effects of non-publishing researchers on the  $P_j$  calculation. According to [6], non-publishing researchers, the potential authors as they put it, are supposed to be considered in the quantification of publication productivity; their population in each

discipline can be estimated with the Waring distribution, which is a probability model used for estimating zero-frequency from zero-truncated data. This could be a potential problem when applying this study's method to the publication data of other research institutions, which cover diverse disciplines of the arts and sciences. To increase the usefulness of the method, the applicability of such a probability model to our study should be examined in the future.

## 4 Conclusions

We proposed a method for the normalization of the number of papers to overcome the difficulties in fair comparison of researcher's publication productivity over diverse disciplines. In our analysis of the publication data of Kyutech indexed in the Web of Science and Scopus, similar variations were found in the calculated values of publication productivity  $P_j$  in the well-populated disciplines; for example, the  $P_j$  values of Engineering and Computer Science were statistically lower than the average in common. The normalization resulted in a relative increase in the total number of papers in disciplines with low publication productivity and a decrease in disciplines with high publication productivity. We further investigated the minimal sample size required for accurate determination of  $P_j$ . When all the Kyutech data were taken as the parent population and bootstrap calculations were made, changing the sample size little by little, it was suggested that the sample size should be greater than 200 for the Web of Science data and greater than 300 for the Scopus data for determining  $P_j$  with accuracy. Because the amount of data did not meet the requirements in every discipline, the confidence intervals of  $P_j$  are considered to have been overestimated in width in the present study. With the normalized index DWPP, compared to the naïve index of number of papers, researchers publishing papers in disciplines with low publication productivity enjoyed a marked rise in ranking. The change of a measure mostly affected the rankings of middle-ranking researchers; the ranking of researchers with extremely high or low productivity were less affected. In summary, the database selection did not matter for the results in the analysis of Kyutech's publication data. Note that the wider coverage of journals in Scopus enabled us to determine the rankings of more researchers.

The above-mentioned wide confidence intervals of  $P_j$  make it difficult to make a precise analysis for its disciplinary differences and creates uncertainty about the resulting normalized index. Therefore, we are planning to collaborate with other research institutions, to collect more publication data. For reliable analysis, name disambiguation of papers is crucial because the total publication effort ratio of researchers should be determined in each discipline. Our final aim is to conduct performance evaluation of research at various levels, such as the level of the whole institution, department, and individual, in a way that is as unbiased as possible, across disciplinary boundaries.

## Acknowledgments

We are grateful to the staff of the General Affairs Division of Kyushu Institute of Technology for their contribution in conducting name disambiguation for papers in the Web of Science. We also thank the staff of the Research Cooperation Division of the institute for their support in collecting the publication data of Elsevier.

## References

- [1] H. Fukunari, "Practical and applied bibliometric analysis for research strategy," *Journal of Information Processing and Management*, 57(6), pp.376–386, 2014 (in Japanese).
- [2] L. Butler, "Assessing university research: a plea for a balanced approach," *Science and Public Policy*, 34(8), pp.565–574, 2007.
- [3] N. Furubayashi, "Beyond bibliometrics. Research analytics and evaluation based on citation data," *Information Science and Technology Association*, 64(12), pp.520–526, 2014 (in Japanese).
- [4] Thomson Reuters, "InCites Indicators Handbook," pp.28, 2014.
- [5] Elsevier, "SciVal Metrics Guidebook," pp.95, 2014.
- [6] T. Koski, E. Sandström, and U. Sandström, "Towards field-adjusted production: Estimating research productivity from a zero-truncated distribution," *Journal of Informetrics*, 10(4), pp.1143–1152, 2016.
- [7] L. Butler, "Explaining Australia's increased share of ISI publications – the effects of a funding formula based on publication counts," *Research Policy*, 32(1), pp.143–155, 2003.
- [8] J. W. Schneider, "Publications or citations – does it matter? Beneficiaries in two different versions of a national bibliometric performance model, an existing publication-based and a suggested citation-based model," *Proceedings of ISSI 2015 Istanbul*, pp.477–488, 2015.
- [9] L. Colledge, "Snowball metrics recipe book," Amsterdam: Snowball Metrics Program Partners, pp.110, 2014.
- [10] Research project for the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, "Survey analysis report on performance evaluation of researchers," 2015 (in Japanese).
- [11] L. R. Vinluan, "Research productivity in education and psychology in the Philippines and comparison with ASEAN countries," *Scientometrics*, 91(1), pp.277–294, 2011.
- [12] J. S. Barrot, "Research impact and productivity of Southeast Asian countries in language and linguistics," *Scientometrics*, pp.1–15, 2016; doi:10.1007/s11192-016-2163-3.
- [13] D. Hicks, "Performance-based university research funding systems," *Research policy*, 41(2), pp.251–261, 2012.
- [14] G. Sivertsen, "Publication-based funding: the Norwegian model," *Research Assessment in the Humanities*, Springer, pp.79–90, 2016.
- [15] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National academy of Sciences of the United States of America*, pp.16569–16572, 2005.

- [16] G. Abramo and C. A. D'Angelo, "How do you define and measure research productivity?," *Scientometrics*, 101(2), pp.1129–1144, 2014.
- [17] A. Schubert, W. Glänzel, and T. Braun, "Against absolute methods. Relative scientometric indicators and relational charts as evaluation tools," In: A. van Raan (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North-Holland Publishing Company, pp.137–176, 1988.
- [18] P. Mongeon and A. Paul-Hus, "The journal coverage of Web of Science and Scopus: a comparative analysis," *Scientometrics*, 106(1), pp.213–228, 2016.
- [19] L. Egghe, R. Rousseau, and G. Van Hooydonk, "Methods for accrediting publications to authors or countries: Consequences for evaluation studies," *Journal of the American Society for Information Science*, 52(2), pp.145–157, 2000.
- [20] J. Wang and Y. Sakurai, "An introduction to bootstrap methods," Kyoritsu Shuppan Co., LTD, pp.236, 2011 (in Japanese).
- [21] R Development Core Team, "R: A language and environment for statistical computing. R Foundation for Statistical Computing," Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>, 2005.