NBER WORKING PAPER SERIES

BIGGER THAN YOU THOUGHT:
CHINA'S CONTRIBUTION TO SCIENTIFIC PUBLICATIONS

Qingnan Xie
Richard B. Freeman

Working Paper 24829
http://www.nber.org/papers/w24829

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2018

We thank participants in the following seminars and conferences for comments on the earlier stages of this paper: China Economic Seminar, Harvard University Department of Economics; Economics of Science and Engineering Workshop, Harvard University Department of Economics joint with the Harvard Business School; Meeting of the International Network on the Value of Health Research, Institute for Fiscal Studies, London; the ASSA-CES Special Session on Innovation, Entrepreneurship and the Chinese Economy, Philadelphia; The Chinese Socio-Economic Development Symposium, Shanghai Academy of Social Sciences; Empirical Social Science Forum, Shaanxi Normal University, Xi'an; and the Conference on New Aspects of Statistics, Financial Econometrics, and Data Science, Stevanovich Center of the University of Chicago. Financial support from the China Scholarship Council funded the Research Fellowship of Qingnan Xie at the Labor and Worklife Program from 2016-2018 The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Bigger Than You Thought: China's Contribution to Scientific Publications
Qingnan Xie and Richard B. Freeman
NBER Working Paper No. 24829
July 2018
JEL No. F0,O0,O3,O31,O33,O34

ABSTRACT

From 2000 to 2016 China increased its scientific publications in the international journals indexed by Scopus to become the largest contributor to global science, accounting for about 23% of journal articles adjusted for the Chinese share of addresses or names on publications. Publications with all-China addresses contributed the most to the increase, followed by cross-country collaborations and papers by Chinese-named researchers outside the country. The same period also saw a huge increase in scientific publications in Chinese language journals not indexed in Scopus. We estimate that while Chinese language papers gain about 1/5th as many citations as non-Chinese (largely English) papers in Scopus they are so numerous that even valued as making 1/5th the contribution of a Scopus paper, China accounts for 36% of global scientific papers defined as Scopus papers and China language equivalent papers and for 37% of citations to those papers. China's move to the forefront of scientific inquiry makes it a key driver of the direction of scientific and technological progress and of the knowledge-based economies of the foreseeable future.

Qingnan Xie
Nanjing University of Science & Technology
and Labor and Worklife Program,
Harvard Law School
2362626753@qq.com

Richard B. Freeman
NBER
1050 Massachusetts Avenue
Cambridge, MA 02138
and NBER
freeman@nber.org

From 2000 to 2016, China jumped from bit player to major contributor of articles in international scientific journals in *physical sciences, engineering and mathematics*. Measured by fractionated addresses indexed by the Scopus bibliometric database, China's share of articles jumped from 4% of those published in 2000 to 18.6% of those published in 2016, topping the US's total.[1] But this does not capture the true magnitude of China's contribution to science. The widely used address metric substantially understates China's rise in prominence by giving China no credit for papers written by Chinese researchers outside the country and published with non-Chinese addresses. And the Scopus database excludes the vast bulk of papers published in Chinese language journals, which while less impactful than papers in international journals, contribute new knowledge and its dissemination to the large population of Chinese researchers. Taking account of these contributions and adjusting Chinese language papers to "Scopus equivalence", we estimate that China was responsible for about 36% of 2016 scientific publications and for 37% of citations of 2013 publications[2] – roughly twice China's share of the world population or world GDP.

This article presents the statistics for these claims. Section one describes how we measure country contributions to papers in international science journals. Section two examines the growth of articles in Chinese language scientific journals not indexed by Scopus and develops an exchange rate between Chinese language articles in those journals and non-Chinese language articles in Scopus to measure China's impact on global science writ large.

1. Measuring national contributions to science publications

The standard measure of a country's contribution to scientific publications credits the country with papers having its single-country address and gives it partial credit for cross country collaborations proportionate to the number of country addresses on the paper,allotting 1/2 credit on multi-country addressed papers to a country with half of the addresses; 1/3rd to a country with one-third of addresses and so on. Splitting credit among countries proportionate to number of addresses potentially understates the contribution of countries with multiple researchers working at a single address compared to those with one researcher per address. To deal with this, we credit a country on a cross-country paper by the ratio of the number of authors with that country's address to total authors. Given the large number of Chinese researchers, this raises China's

---

[1] Measured in the Scopus data-base of scientific publications https://www.scopus.com. Reported by National Science Board (2018), Appendix table 5-27 shows China share exceeded 17.8% for US addresses.

[2] We use 2013 as the latest year for citation counts to allow for 3 years' citations through 2016.

estimated contribution modestly.

The big weakness of the address metric is not, however, that it ignores the number of authors at particular country addresses but that it gives *no* credit to a country for the work of its researchers published at non-country addresses. It counts a paper with, say, five Chinese named authors working in the US as a US paper just as it would a paper with five authors having American names. Indicative of the size of the missing contribution, we estimate that 17% of *non-China addressed* articles in 2016 had at least one Chinese named author.[3]

Our analysis uses the following formula to divide country credit for a paper with N authors at A addresses by national background as well as by their address:

(1) Country C contribution of paper i = $\alpha(Ac/A) + (1-\alpha)(Nc/N)$,

where Ac authors have addresses from C (China in our case),[4] and Nc authors have names associated with that country (i.e. Chinese names); and where α is the weight given to addresses vs names, varying from 1 (only addresses matter) to 0 (only names matter). A paper based on research at a unique facility, say the CERN Hadron Collider, would presumably deserve a higher α than a paper by theorists collaborating over the Internet. For simplicity, we give equal weight to addresses and names. Because China's share of addresses and names increased substantially, the choice of α only modestly affect our measures of China's growing contribution to science (Supplementary Material, Appendix A).

Figure 1 displays our estimates of the weighted fractional *contribution* of China to Scopus papers in academic journals based on authors' address and name.[5] We attribute 23.3% of the papers published in 2016 to China, which is 5.3 points higher than the 18.0% given by the standard country weighted address measure of contribution. The figure differentiates three types of Chinese papers: those with China-only addresses; international collaborations with China and non-China addresses; and papers with Chinese-named authors but no China addresses. The largest increase in China's share came from the all-China address papers, which rose from 4.04%

---

[3] Estimated from 20,000 randomly chosen articles in the 2016 Scopus, with persons from mainland or other Chinese speaking areas differentiated from Chinese born outside elsewhere by their first name (Wei is Chinese; James is not) as well as by second name.

[4] We treat authors with multiple institutional addresses in different countries by dividing their contribution to addresses proportionately to the number of addresses by country. If one author on a 2-author article listed one institution in country C and another in country D, we credit those countries with 1/4th from that author.

[5] Our counts are for journal articles only and thus differ from counts in NSB's Science and Engineering Indicators 2018, which cover publications beyond journals.

of all papers in 2000 to 17.87% of all papers in 2016.[6] The proportion of international collaborations went from 0.38% of papers in 2000 to 2.82% in 2016 while the share of papers having Chinese names with no Chinese addresses went up from 2.94% to 5.28%.

Taking the entire 2000-2016 period, China added approximately 2.2 million non-Chinese language papers (weighted by the Chinese share of addresses or names) to the stock of non-Chinese language papers and 1.1 million papers in Chinese language journals indexed in Scopus.

Figure 2 shows China's increased role in the scientific literature from a different perspective – in terms of the proportion of papers with *any association* to China. The association measure is 1 if a paper has at least one Chinese address **or** Chinese named author and 0 otherwise. The association measure exceeds a country's weighted proportion of articles to the extent that researchers collaborate widely outside their country or group. China was associated with 12.4% of papers published in 2000 and 34.5% of papers published in 2016.

All told, Figures 1 and 2 show that China's representation in international scientific journals increased massively in the early 21st century – at rates beyond what knowledgeable experts seemingly viewed as possible as recently as 2011.[7]

Contribution measured by citations

To the extent that researchers pay less attention to papers with Chinese addresses or names than to other papers, the numbers in Figures 1 and 2 exaggerate the increased Chinese contribution to science. The standard measure of the attention given to a paper is the number of citations it receives, which varies for reasons of citation homophily – the tendency for researchers to disproportionately cite researchers with characteristics like themselves; for its "intrinsic quality", and for the citation practices and trajectory of publications in its field.

To measure China's contribution in citations, we counted citations made by all Scopus publications to journal articles published in 2000 and 2013, using a three-year window every year from 2000 to 2013. We estimate the number of citations to papers with China addresses/names relative to all papers and the share of world citations received by those papers.

[6] The 2004-2005 expanded Scopus coverage of Chinese language journals contributed to this, but the vast bulk is through increased publications in non-Chinese language journals. The number of China address published in a non-Chinese language journal increased by 539.2% from 2000 to 2016 compared to a 158.4% increase in Chinese language journals. In 2000 39.1% of China addressed articles were in the Chinese language.

[7] May (1996) placed China between Denmark and Switzerland in publications; Zhou and Leydesdorff (2006) noted China's advance to between France and Germany. Kumar and Asheulova (2011) projected that China would have substantially fewer publications in Scopus than the US through 2025.

Panel A of Figure 3 shows the ratio of average number of citations to Chinese papers to the global average of citations. In 2000 papers with all-Chinese addresses received just 29% of the world average of citations, implying that those papers had little impact on research worldwide. By contrast, papers by Chinese researchers working outside the country or collaborating with researchers overseas received citations above world average. Weighted by the distribution of papers among the groups, Chinese researchers received about the global average in citations. The data show a different picture in 2013. Citations to papers with all-China addresses increased to 70% of the global average[8] Citations to international collaborative papers also increased relative to the global average while citations of papers by Chinese researchers at non-China addresses converged toward the global mean. The huge increase in the all-China addressed papers put the average citation of a 2013 Chinese paper below the world average.

Panel B 3 turns to China's share of global citations – the multiplicand of the relative number of citations and China's share of papers. China's share of global citations rose from 7.4% in 2000 to 19.5% in 2013, due primarily to the increased share of citations to all China addressed papers, which signals a shift in the locus of global science to China.

As a final look at China's growing presence in the international scientific literature, we examined the percentage of Chinese names and Chinese addresses in papers published in Nature and Science. Figure 4 shows that in 2016 about 20% of names on papers in these journals were Chinese while 8% to 9% of addresses were Chinese – far above the percentage Chinese of names or addresses in 2000, with the bigger increase occurring in addresses.[9]

In sum, all of our measures of China's contribution to the scientific literature show that China increased its scientific contribution to levels far above what one might expect from the country's share of world population or world national production.[10]

2. Missing matter: Chinese language papers and citations

The expansion of Chinese publications in primarily English Scopus journals could

---

[8] Xie and Freeman (July 2018) show this reflects both the increased number of Chinese papers, which boosted citations to Chinese papers due to citation homophily and improved quality reflected in a rising trend of citations to Chinese papers from papers with no China connection..

[9] Bornmann, Leydesdorff, and Wagner (2015) show an increase in BRICS country presence on highly cited papers and a strong China connection with US; Wang, Wang and Philipsen (2017) describe growing co-authorship between Chinese based and EU researchers

[10] In 2018 China had 18.2% of world population (https://en.wikipedia.org/wiki/List_of_countries_and_dependencies _by population). In 2017 it had 18.3% of world GDP in purchasing power parity terms. (https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP) population.

reasonably be expected to have come at the expense of publications of Chinese language scientific articles. Historically, the spread of English as the language of science in the 20th century reduced the number of journals and articles in other languages.[11] To our surprise, this was not the case for China. The number of Chinese language scientific papers in the comprehensive China National Knowledge Infrastructure (CNKI) [12] database of journals published in China shows a growing number of papers in the 2000-2016 period with a level that reveals a huge "missing matter" invisible science problem[13] (Wagner and Wong, 2012; Van Leeuwen et al 2001) in using Scopus to assess the level and growth of China's contribution to global science. In 2007 the CNKI listed 4,216 science, engineering, and math journals as of 2017[14] compared to 329 active Chinese language science journals in Scopus. While Scopus presumably chose the most outstanding Chinese journals for inclusion, the 92% of Chinese journals that it did not cover almost certainly had to contain some useful and interesting science.

Figure 5 records the number of CNKI articles in sciences, engineering, and math journals from 1980 through 2016; the number less the overlap in journals indexed by Scopus; and the total number of Scopus articles. In the 1980s the number of CNKI articles fell far short of the number of Scopus articles. China's contribution to science was modest because its low GDP per capita precluded supporting much scientific work and because the 1966-76 Cultural Revolution had destroyed the country's university system and curtailed research. In the 1990s China rebuilt its higher education system sufficiently rapidly to increase the number of articles in the CNKI to rough parity with the number in Scopus. Thereafter the number of CNKI articles increased more and then less rapidly than Scopus articles so that in 2016 the two databases published approximately the same number of articles (1.6 million).

To gain greater insight into the relation between CNKI and Scopus aticles, we compared the number of Scopus articles with China addresses with CNKI articles in 12 detailed fields[15]

---

[11] Gordin (2015).

[12] The CNKI includes PhD dissertations, masters' theses, proceedings, newspapers, yearbooks, yearbooks, e-books, patents, and covers humanities, arts, economics, and business as well as physical sciences, engineering, and mathematics.

[13] We randomly sampled 10,000 CNKI Chinese language articles published in 2016 and found all articles had at least one China addresses and 9,957 articles had only Chinese names.

[14] In addition, the CNKI reports 191 non-Chinese language journals published in China, largely in English. In 2016, 95.67% of CNKI scientific journals were in Chinese, 4.06 % are in English, and 0.27% in other languages.

[15] Because CNKI definitions of fields are closer to those in the Web of Science than to field definitions in Scopus, this analysis used journal articles from Web of Science (which has a similar English language journal coverage to Scopus).

and in universities of different status. CNKI Publications fell in mathematics, optics, metallurgy, and instrumentation; held steady in microbiology, and increased in seven fields such as oncology and pediatrics (Supplementary Material, Appendix B). Among universities, researchers at the highest quality “985” universities reduced their CNKI publications while increasing their English language papers; but the growing number of researchers in less prestigious universities maintained or increased CNKI publications, possibly filling space when top researchers shifted to international journals, while also publishing in English language journals (Supplementary Material, Appendix C). Some scientists may also have published similar articles in English to reach global researchers in the global community and in Chinese to spread findings to potential users in China – for instance, biomedical researchers informing Chinese doctors about findings relevant to medical practices.

The underlying reason for China's concordant increase in publications in international Scopus journals and CNKI journals appears, however, to be the huge increase in the number of researchers seeking publication of research findings fueled by the country's massive investment in R&D, in university faculty, and in industrial researchers. Between 2000 and 2016 China increased its R&D spending tenfold in constant purchasing power parity terms to exceed EU spending and approach that in the US.[16] Over the same period China more than doubled its number of faculty and tripled its number of researchers – all of whom had to find venues for publishing.[17]

Quality and impact

If the scientific content/impact of Chinese language papers was on a par with that of English language papers, the roughly equal number of CNKI and Scopus articles published in 2016 would imply that China was responsible for about 2/3rds of scientific work (½ from CNKI journals and ~ 1/3 of the ½ from Scopus) in that year! But CNKI articles do not have the same quality/impact as international journal articles. Fewer scientists read Chinese than English.[18]

[16] The 2000 to 2016 growth figures are from https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm, measured in US dollars in constant prices using 2010 base year and Purchasing Power Parities. The NSB’s Science and Engineering Indicators 2018, Table 4-5 shows that in 2015 China exceeded the EU in purchasing power parity dollars spent on R&D and was sufficiently close to US spending to likely exceed that by 2020, at the latest.

[17] The China Statistical Yearbook shows a 146.2% increase in number of faculty from 2000 to 2014 and a 302.5% increase in the number of researchers. (The China Statistical Yearbook 2001-2015, Table 20-22 and Table 18 respectively). Xie, Zhang and Lai (2014) document China's expansion in higher education and researchers.

[18] Fung (2008) describes issues of English authors/journal citation of Chinese language articles. Panko (2017) notes problems with having a single dominant scientific language.

CNKI papers are shorter and have fewer references than Scopus papers and thus presumably encapsulate less knowledge.[19] China's requirement that PhD's and master's candidates publish parts of their thesis to obtain a degree leads to many publications with a narrow scope that would place them in university libraries rather than scientific journals elsewhere. Indicative of the quality difference, 44.6% of CNKI papers published in 2013 received no citations compared to 29.0% of Scopus papers published in 2013.[20] Recognizing the higher impact/quality of English language publications, Chinese universities offer sizable monetary and promotion incentives for publishing in those journals,[21] which induces many researchers to send their best work overseas.

Stipulating that CNKI publications have lower quality/impact than Scopus publications, the key issue in aggregating the numbers of articles in the data sets relates to their exchange rate: How valuable is a CNKI article relative to a Scopus article in contributing to science?

We answer this question by examining citations within the Scopus and CNKI databases and between them. Both Scopus and the CNKI count citations within their own database. In 2013 a Scopus journal article averaged 9.2 forward citations from other Scopus articles while a CNKI journal article averaged 2.3 forward citations from other CNKI articles.[22] This suggests a citation-based exchange rate of about 0.25 (= 2.3/9.2) for CNKI articles compared to Scopus articles.

As neither database counts citations from the other, we estimated cross-database citations from the referencesvto articles published in 2013 in random samples of non-Chinese language Scopus papers and of Chinese language papers in CNKI.[23] Our calculations revealed a huge imbalance in citation flows. CNKI articles cited many articles in Scopus while Scopus articles cited few CNKI articles. We estimate that about half of journal references in CNKI articles to articles published in 2013 were to Scopus articles, while just 0.29% of journal references in

---

[19] We documented the samples of 2,000 randomly selected CNKI journal articles and found nine references per article compared to 42 references per article in Scopus. To the extent that articles with fewer references rely on less information and cover less material than articles with a greater number of references, a CNKI article has less scientific value than Scopus article, possibly by the 9/42 ratio references and thus to be about 21% as informative.

[20] These estimates based on all journal articles in Scopus and the CNKI from August 2017 to November 2017

[21] Quan, Chen and Shu (2017).

[22] The distribution of citations in both data sets follow power laws with most papers receiving few citations while a few gain lots, with Scopus papers have a heavier tail than CNKI papers. Correlatingcitations from Scopus and from CNKI in the overlap Chinese language journals covered by both we found a significant positive correlation.

[23] We randomly sampled 10,000 non-Chinese language articles in Scopus from 2013 to 2017 and found 19,859 references to journal articles published in 2013. We randomly sampled 2,000 Scopus non-Chinese language articles from 2013 to 2017 to obtain their references and randomly sampled 500 articles from CNKI Chinese language articles to obtain their references . Supplementary Materials, Appendix D gives the details of our estimation

Scopus articles to 2013 published articles were to Chinese language journals in the CNKI and not in Scopus. (Supplementary Materials: Appendix D). Extrapolating these estimates to all 2013 articles we estimate that 2013 Scopus articles received 3,276,350 citations from CNKI articles in the following three years whereas 2013 CNKI articles received l32,196 citations from non-Chinese language Scopus articles. Adding these citations to the number of citations in Scopus and CNKI reduces the exchange rate of a CNKI journal article from 0.25 to 0.20 of a Scopus paper.

Figure 6 shows that with this adjustment, China contributed 36% of Scopus equivalent "global science" articles based on fraction weighted names and addresses and gained 37% of Scopus equivalent citations while having an association with 45% of Scopus equivalent articles. As our calculations ignore scientific articles outside Scopus in languages other than Chinese, this over-estimates China's share of "global publications". But because Scopus includes many non-English non-Chinese journals (14% of Scopus journals are in non-English non-Chinese journals) and because no country comes close to China in its scale of science, such adjustments will not substantively alter our finding that China's expansion in science is "bigger than you thought".

3. Conclusion

That China, one of the lowest income countries in the world at the turn of the 21st century, became a super-power in scientific knowledge in less than two decades is a remarkable development in the history of science. The way China deploys its newly developed scientific resources will help drive the direction of science and technology into the foreseeable future; and given the role of scientific and engineering knowledge in modern economies and society, give the country a huge role in developing the global knowledge-based economy. To paraphrase Horace Greeley's advice to Americans as the US expanded to California "Go West, young man, and grow up with the country,"[24] barring some huge massive change, science is going East and will grow up with China.

[24] https://www.encyclopedia.com/history/dictionaries-thesauruses-pictures-and-press-releases/go-west-young-man-go-west

References

Bornmann, Lutz, Caroline Wagner, and, Loet Leydesdorff. 2015. “BRICS Countries and Scientific Excellence: A Bibliometric Analysis of Most Frequently Cited Papers,” *Journal of the Association for Information Science and Technology* 66(7) (July): 1507–1513.

China National Knowledge Infrastructure, http://www.cnki.net/ net

National Bureau of Statistics, PRC. For the years 2000-2014. *China Statistical Yearbook*. China Statistics Press, 2000-2014.

Freeman, Richard B., and Wei Huang. 2014. "Strength in Diversity," *Nature* 513(7518) (September 16): 305. Available at: https://www.nature.com/news/collaboration-strength-in-diversity-1.15912

Freeman, Richard B., Ina Ganguli, and Raviv Murciano-Goroff. 2015. “Why and Wherefore of Increased Scientific Collaboration, in Adam B. Jaffee and Benjamin F. Jones (eds) in *The Changing Frontier: Rethinking Science and Innovation Policy* (Chicago, IL: University of Chicago Press for NBER): 17-48.

Fung, Isaac C.H. 2008. “Citation of Non-English Peer Review Publications – Some Chinese Examples,” *Emerging Themes in Epidemiology*, 5:12. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570362/

Gordin, Michael D. 2015. *Scientific Babel: How Science Was Done Before and After Global English* (Chicago, IL: University of Chicago Press).

Kumar, Naresh, and Nadia Asheulova. 2011. “Comparative Analysis of Scientific Output of BRIC Countries. *Annals of Library and Information Studies* 58(3) (September): 228-236. Available at: http://nopr.niscair.res.in/bitstream/123456789/12846/1/ALIS%2058(3)%20228-236.pdf

Leydesdorff, Loet, Carolina S. Wagner, Lutz Bornmann. 2014. “The European Union, China, and the United States in the Top-1% and Top-10% Layers of Most-Frequently Cited Publications: Competition and Collaborations,” *Journal of Informetrics* 8(3) (July): 606-617. Available at: https://arxiv.org/abs/1404.0462

May, Robert M. 1997. “Scientific Wealth of Nations,” *Science* 275(5301) (February 7): 793-796.

National Science Board. 2018. *Science and Engineering Indicators 2018*. NSB-2018-1. Alexandria, VA: National Science Foundation. Available at: https://www.nsf.gov/statistics/2018/nsb20181/

Panko, Ben. 2017. “English Is the Language of Science. That Isn't Always a Good Thing,” *Smithsonian* (January 2). Available at: https://www.smithsonianmag.com/science-nature/english-language-science-can-cause-problems-180961623/

Zhou, Ping, and Loet Leydesdorff. 2006. “The Emergence of China as a Leading Nation in Science,” *Research Policy* 35(1): 83-104. Available at: https://arxiv.org/abs/0911.3421

Quan, Wei, Bikun Chen, and Fei Shu. 2017. "Publish or Impoverish: An Investigation of the Monetary Reward System of Science in China (1999-2016)," Aslib Journal of Information Management, 69(5): 1-18. Available at: https://arxiv.org/abs/1707.01162
http://www.sciencemag.org/news/2017/08/cash-bonuses-peer-reviewed-papers-go-global

Scopus Data Base. https://www.scopus.com

Van Leeuwen, Thed N., Henk F. Moed, Robert J.W. Tijssen, Martijn S. Visser, and Anthony F.J. Van Raan. 2001. “Language Biases in the Coverage of the Science Citation Index and its Consequences for International Comparisons of National Research Performance,” *Scientometrics* 51(1) (April): 335–346.

Wagner, Caroline S., and Shing Kit Wong. 2012. “Unseen Science: Representation of BRICs in Global Science,” *Scientometrics* 90(3) (March): 1001-1013.

Wikipedia contributors. 2018. "World population," *Wikipedia, The Free Encyclopedia*, https://en.wikipedia.org/w/index.php?title=World_population&oldid=848888642 (accessed July 7, 2018).

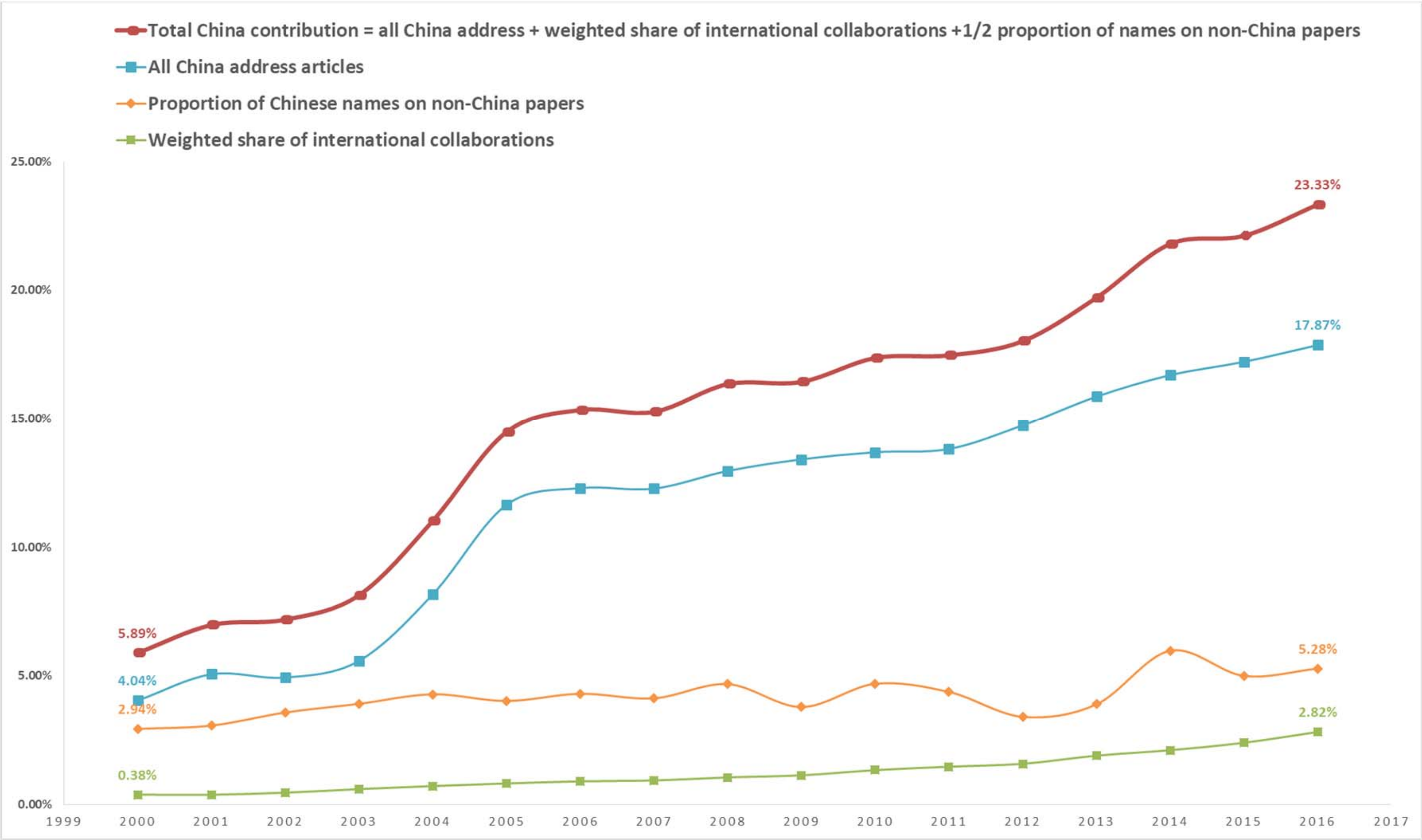
World Bank Data. https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.CD

Wang, Lili, Xianwen Wang, and Niels J. Philipsen. 2017. “Network Structure of Scientific Collaborations Between China and the EU Member States,” *Scientometrics* 113(2): 765–781.

Xie, Qingan, and Richard Freeman. 2018. “Improved Quality or Citation Homophily? the Increase in Citations to Developing Country Research” (NBER WP, July 2018)

Xie, Yu, Chunni Zhang, and Qing Lai. 2014. “China’s Rise as a Major Contributor to Science and Technology,” *PNAS* 111(26) (July): 9437-9442.

Figure 1: Weighted Share of International Journal Articles to China, 2000-2016



Note: Data calculated from Scopus data base classified by their year of publication. Proportion of articles with non-Chinese addresses but at least one Chinese name estimated from random sample of 20,000 Scopus articles with non-Chinese addresses in each year.

Figure 2: China-associated Articles in Scopus

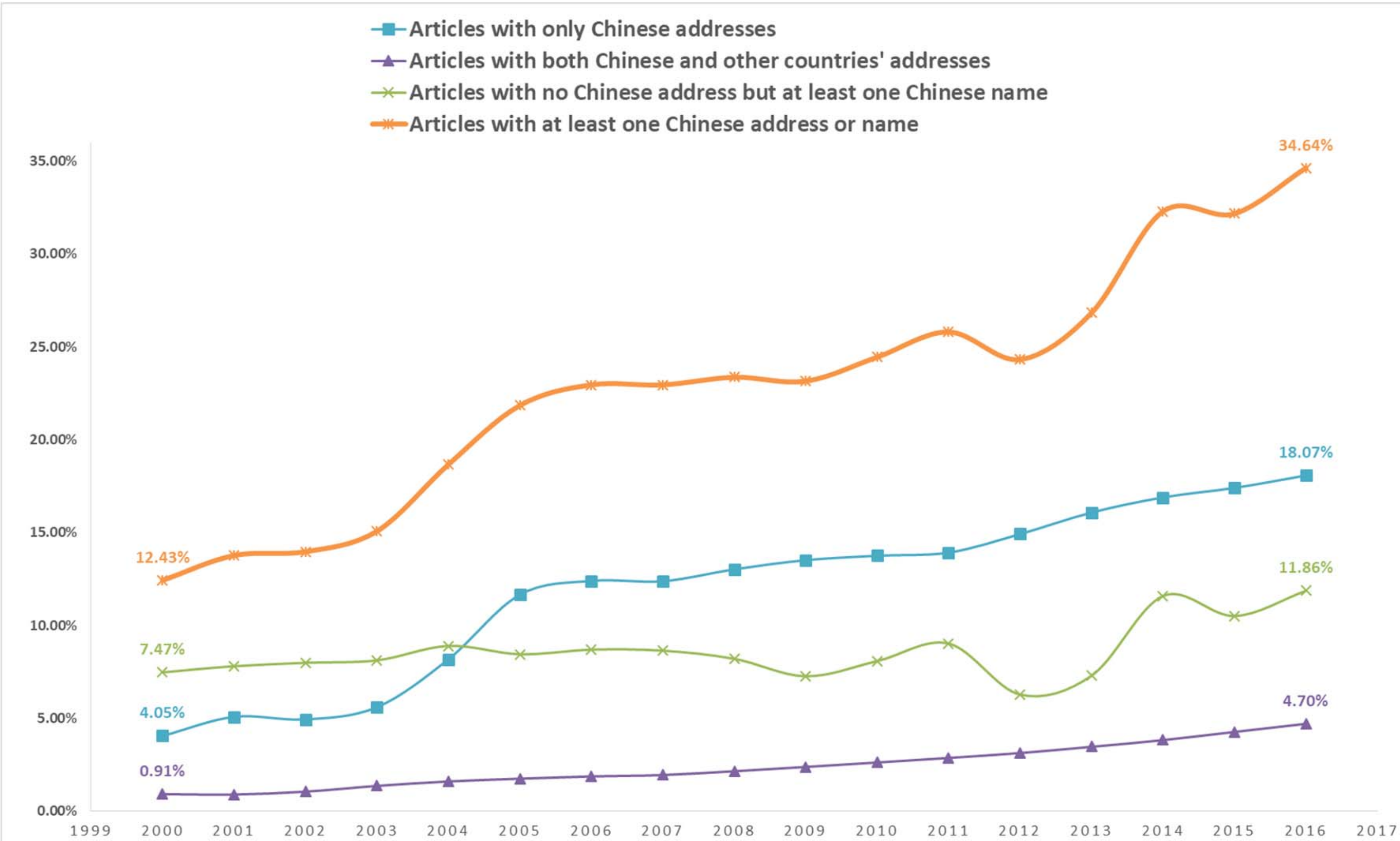


Figure 3: Ratio of Average Citations of China-linked papers to Average World Citations and Share of China-Linked Papers in World Citations in three year interval, 2000 and 2013

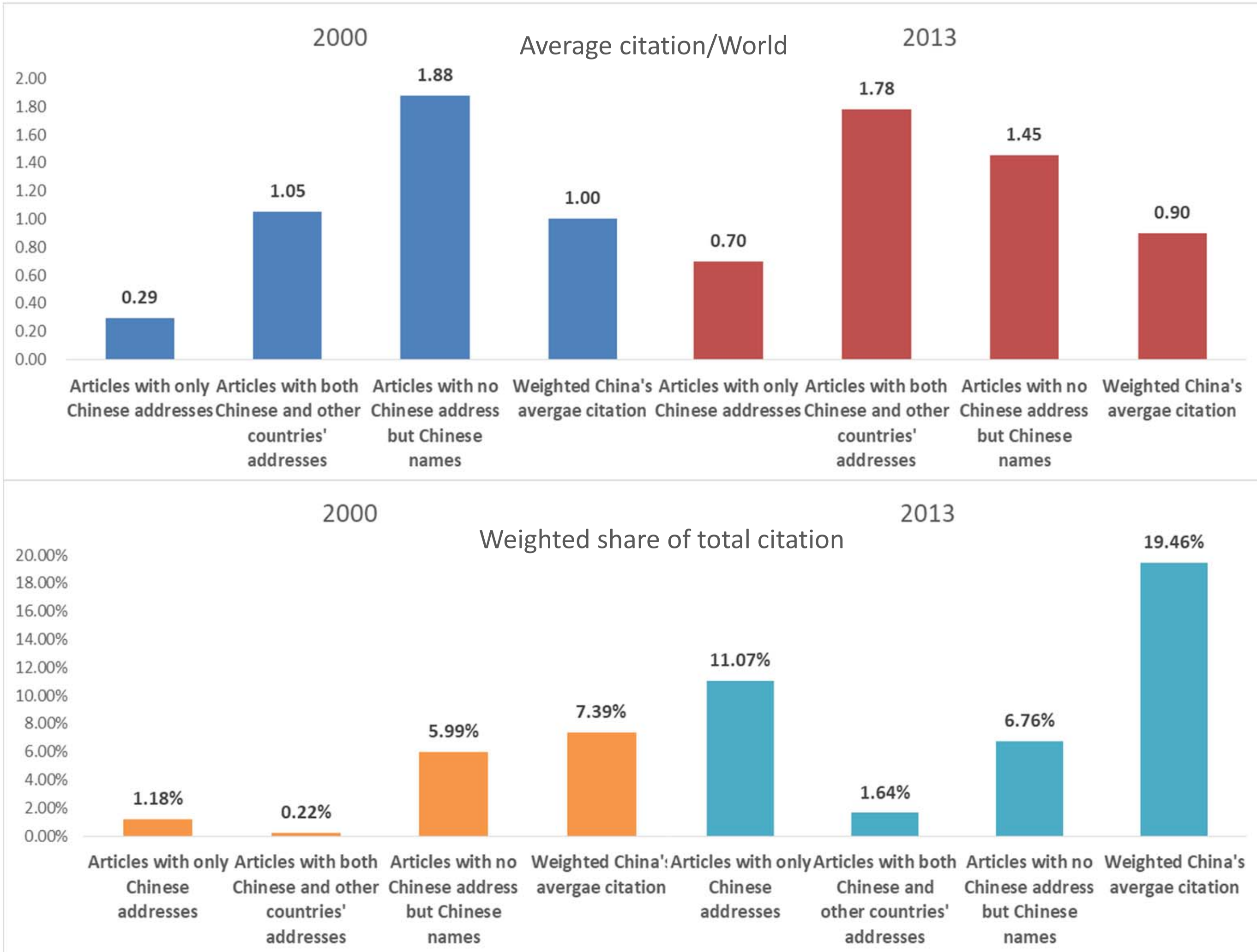


Figure 4: Fraction Weighted Chinese Addresses and Names on Articles in Nature and Science, 2000 and 2016

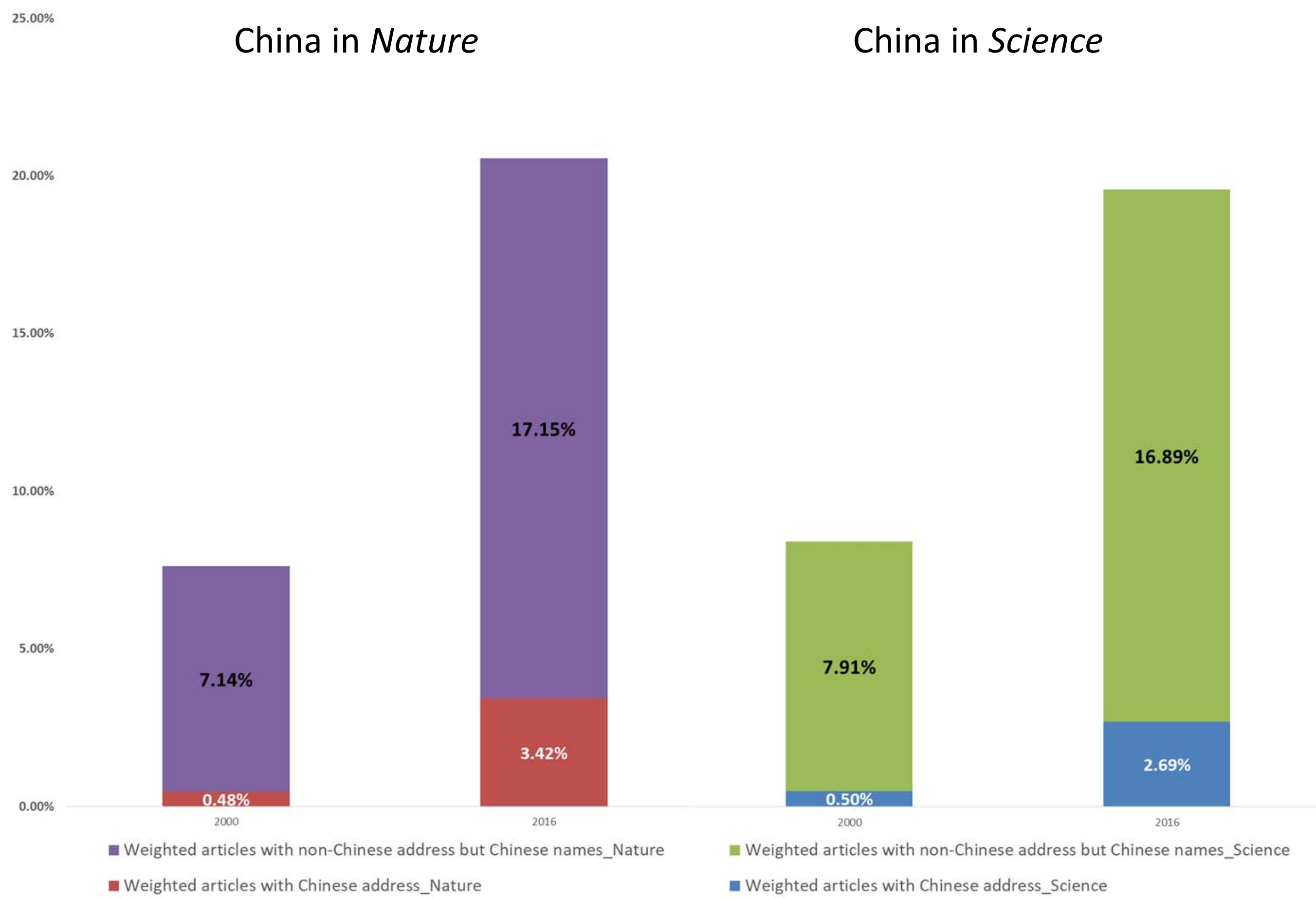


Figure 5: Numbers of Science, Engineering, and Math Journal Articles in CNKI and Scopus, 1980-2016

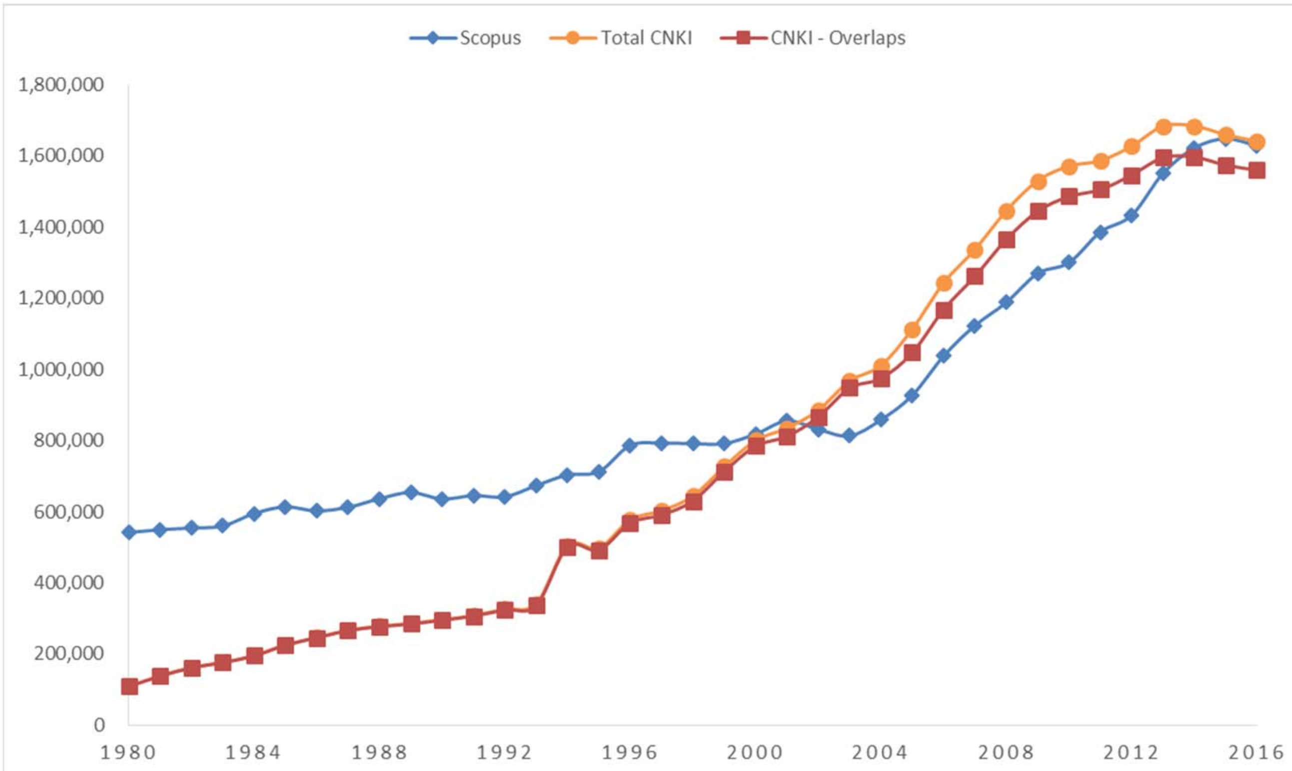
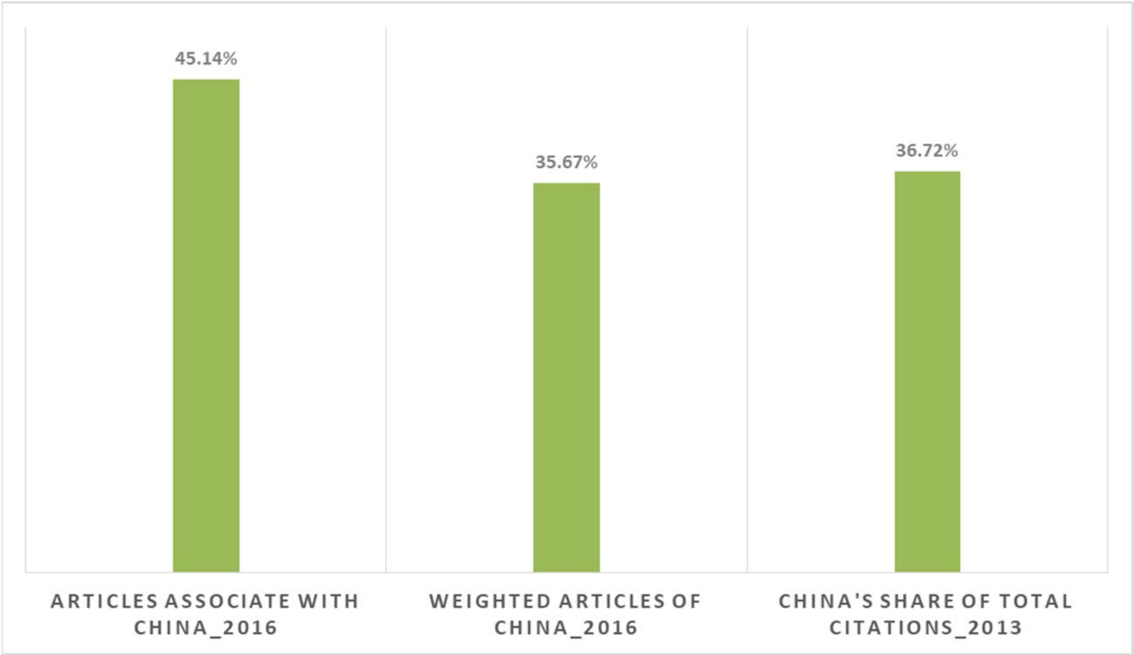


Figure 6: China's Share of 2016 Global Science Publications and 2013 Global Citations



Note: Based on estimated Scopus equivalence of CNKI articles as ratio of average citations of CNKI articles/average citations of Scopus articles = 0.20 in our calculations.

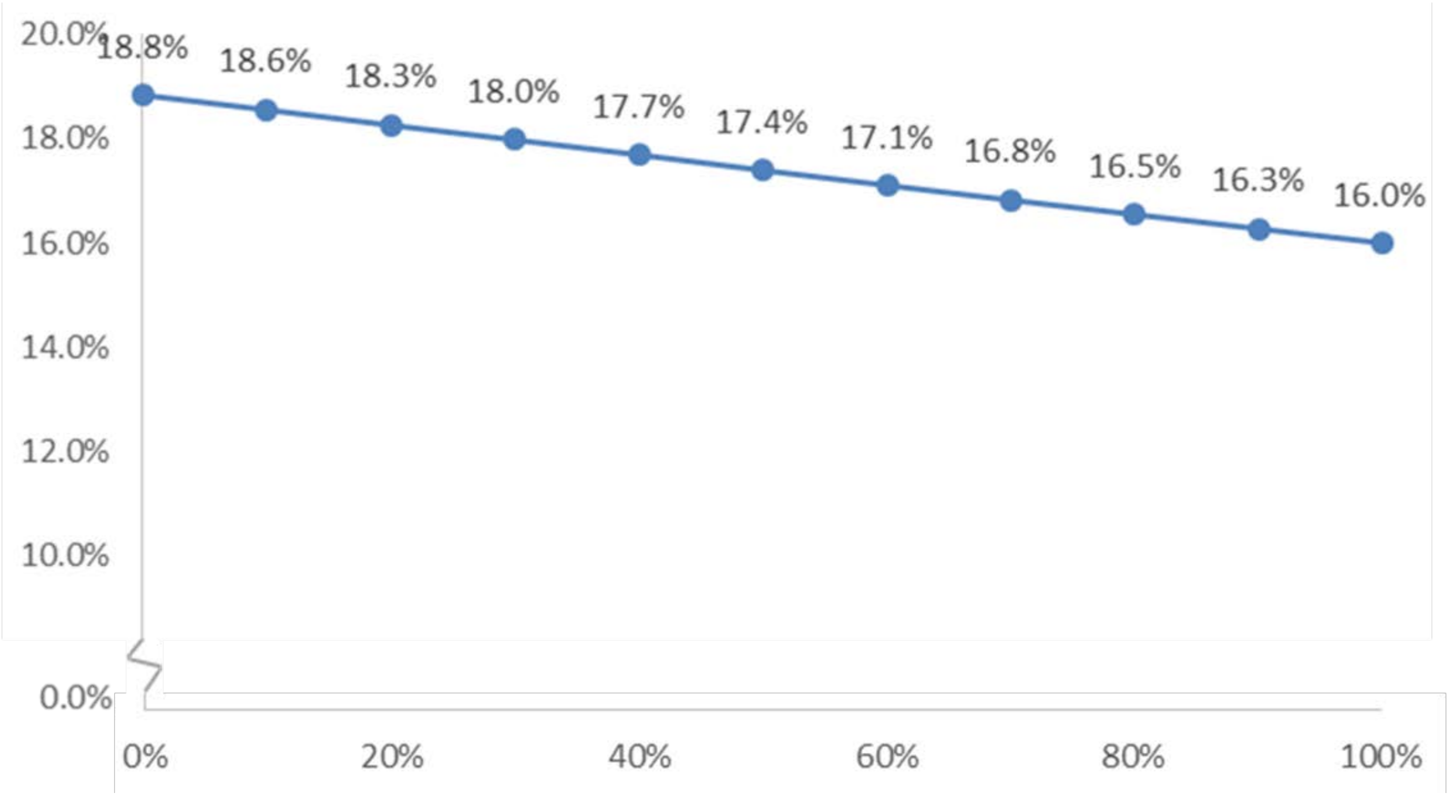
Share of articles associated with China = (# China associated articles in Scopus + 0.20 * # CNKI only Chinese articles) /(# articles in Scopus + 0.20 * # CNKI only Chinese articles).

Share of weighted articles of China = (# weighted Scopus articles of China + 0.20 * # weighted CNKI only Chinese articles)/ /(# articles in Scopus + 0.20 * # CNKI only Chinese articles).

Share of citations = (citations on Scopus paper weighted by China names/addresses + weighted citations of China from CNKI only Chinese articles to Scopus non-Chinese language articles + total CNKI citations from Scopus and CNKI)/total global citation)

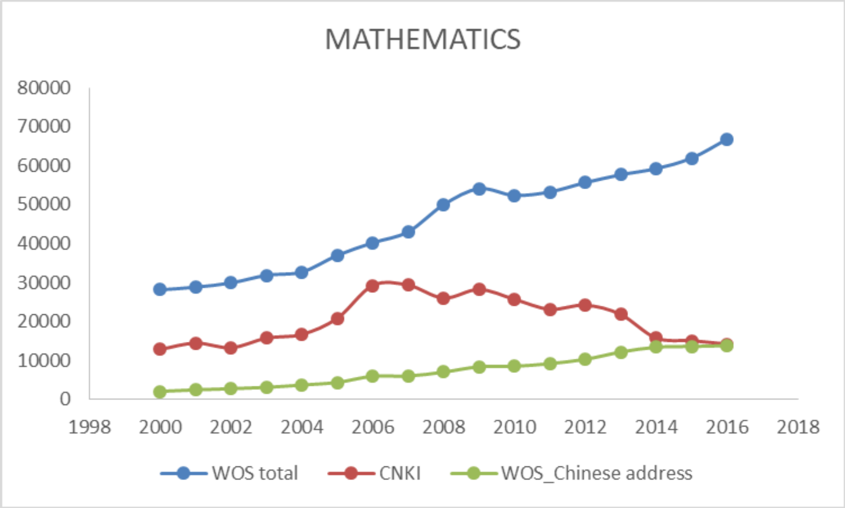
Appendix-Supplementary material

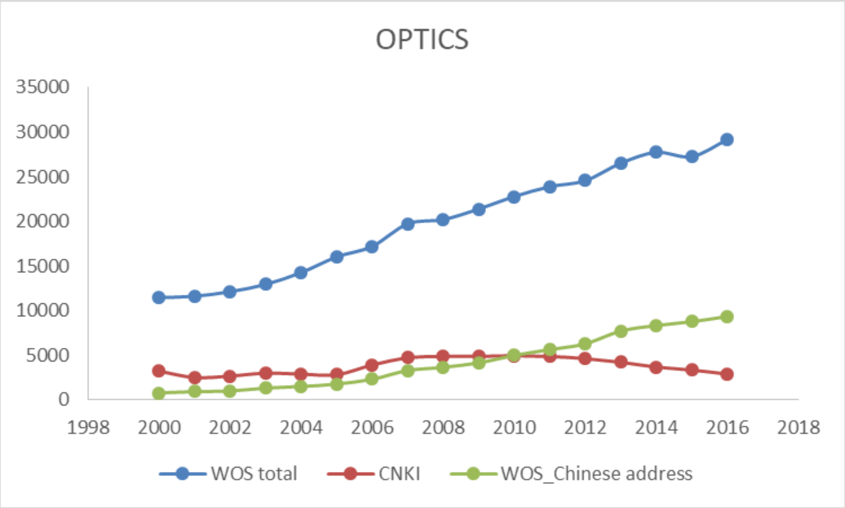
Appendix A: China Share of Articles With Different α weights on Addresses

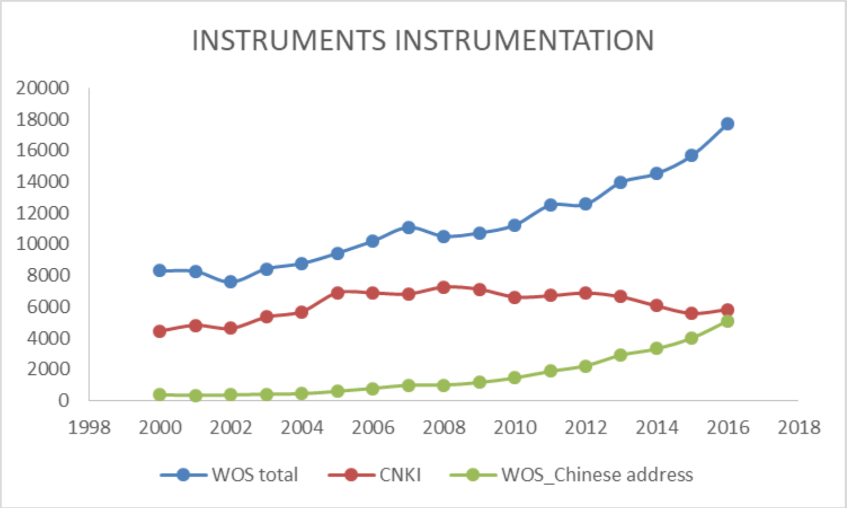


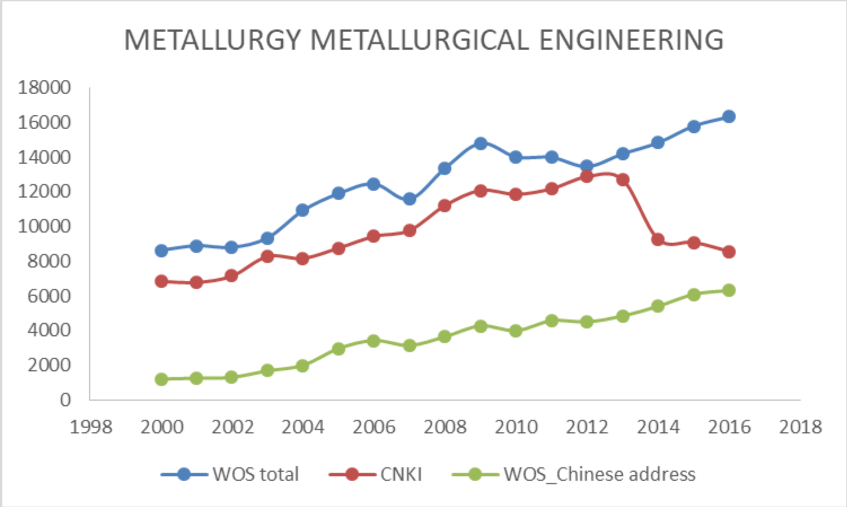
Note: Exhibit A captures the growth of China's weighted share of Scopus journal articles between year 2000 and 2016 changes with α ranges from 0% to 100%.

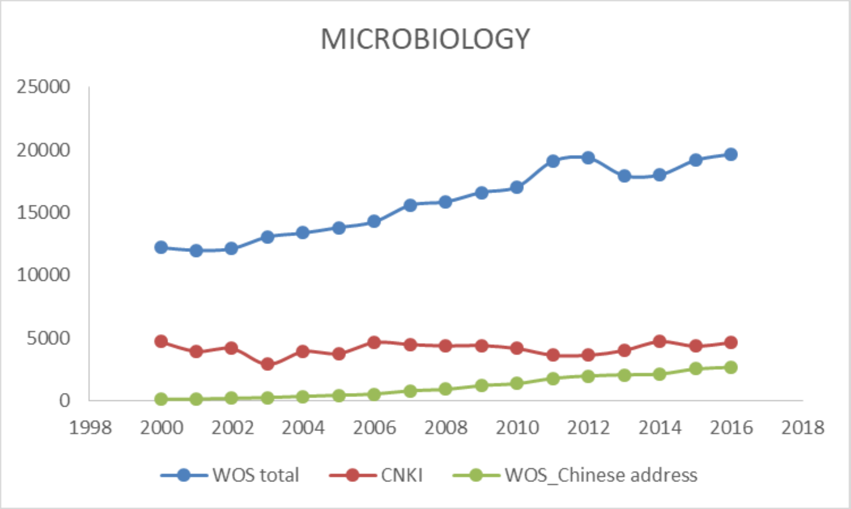
Appendix B: Number of CNKI and Scopus Articles in 12 Fields

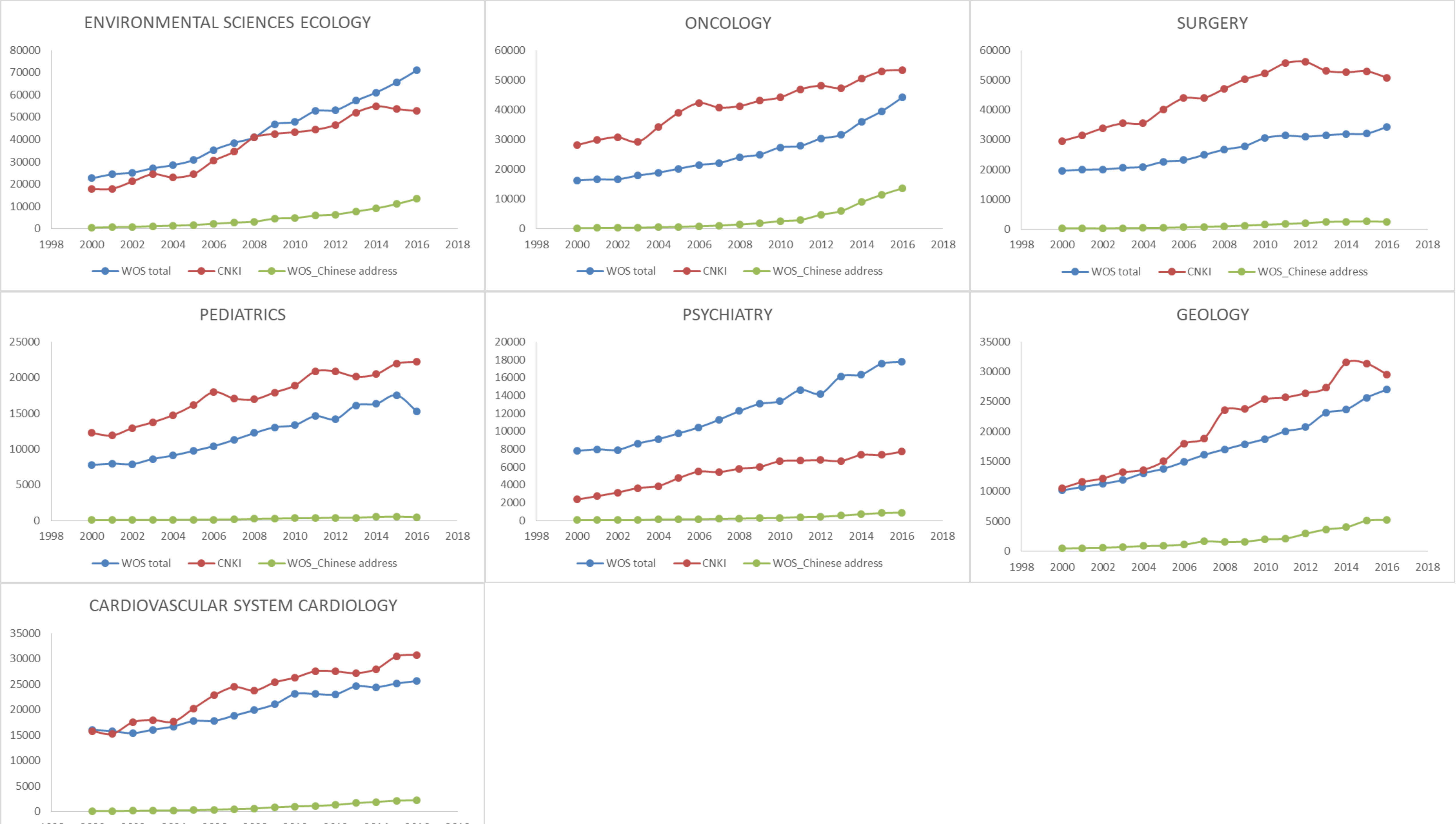






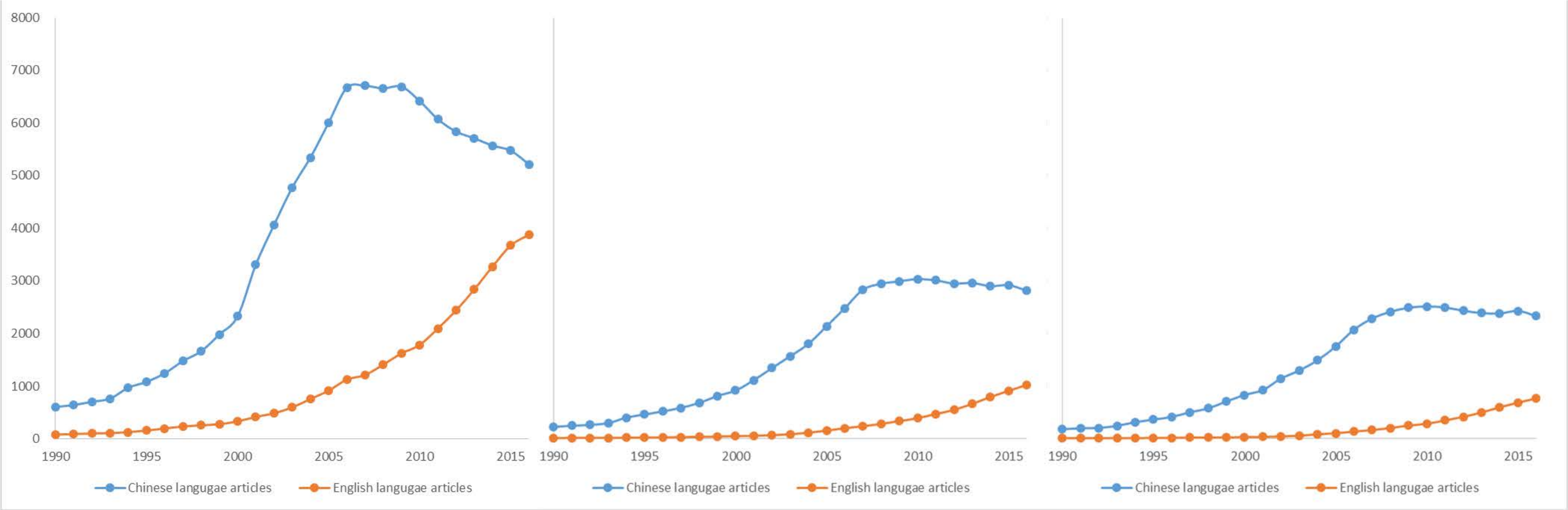





ENVIRONMENTAL SCIENCES ECOLOGY
ONCOLOGY
SURGERY
PEDIATRICS
PSYCHIATRY
GEOLOGY
CARDIOVASCULAR SYSTEM CARDIOLOGY
WOS total
CNKI
WOS_Chinese address

Appendix C: Average Number of Chinese and English Language Articles in Three Tiers of Chinese Universities, 1990-2016

985 Universities: First Tiers | 211 Universities: Second Tiers | Top 50 other Universities: Third Tiers



Appendix D: Estimated Citations to Scopus and CNKI Journal Articles, including Cross-database Citations

	Number of total citations	Average citation per article
Total Scopus	17,533,029	11.31
Scopus to Scopus	14,256,679	9.19
CNKI only Chinese articles to Scopus non-Chinese articles	3,276,350	
Total CNKI	3,831,190	2.28
CNKI to CNKI	3,798,994	2.26
Scopus non-Chinese articles to CNKI only Chinese articles	32,196	

Source: We estimated the number of references that Scopus non-Chinese language articles give to Chinese language journals in CNKI but not in Scopus in 2013 by sampling 10,000 articles from 2013 to 2017 (2,000 per year) and counted the number of journal articles that referenced articles published in 2013. We found 19,859 journal references, nearly all (19,731) to Scopus journals and 64 to Scopus Chinese journals. We selected references with a journal title from the remaining references and matched journal titles with CNKI journals and found 58 references.

We estimated the number of references from CNKI Chinese language articles to Scopus non-Chinese language articles in 2013 by randomly sampling 500 articles from 2013 to 2017 (100 yearly) and found references to 2,984 documents -- 1,031 of which are Chinese language journal articles and 1,848 to non-Chinese language documents. Of these 533 had the mark [J] that CNKI uses to identify journals. But we also found 534 articles that we identified as in journals, giving us a total of 1,067 references to non-Chinese journals. Thus we estimate that 50.87% (= 1067/(1067+1031)) of CNKI references were to non-Chinese journals nearly all in Scopus. A similar analysis of Chinese language papers in Scopus journals gave an estimate of 49.7% of journal citations to non-Chinese Scopus articles. The weighted Chinese address/authors' contribution to the 1067 CNKI references was 37.84% – nearly double China’s 19.46% share of total citations, reflecting homophily in references.

Appendix data
Appendix E: Data for Figure 1

	Total China contribution = all China address + weighted share of international collaborations +1/2 proportion of names on non-China papers	All China address articles	Proportion of Chinese names on non-China papers	Weighted share of international collaborations
2000	5.89%	4.04%	2.94%	0.38%
2001	6.97%	5.06%	3.07%	0.38%
2002	7.18%	4.93%	3.57%	0.46%
2003	8.13%	5.58%	3.91%	0.60%
2004	11.02%	8.16%	4.28%	0.72%
2005	14.47%	11.64%	4.02%	0.82%
2006	15.34%	12.29%	4.30%	0.90%
2007	15.27%	12.27%	4.13%	0.93%
2008	16.35%	12.96%	4.68%	1.05%
2009	16.44%	13.41%	3.79%	1.13%
2010	17.37%	13.69%	4.69%	1.34%
2011	17.47%	13.82%	4.37%	1.47%
2012	18.02%	14.74%	3.40%	1.58%
2013	19.71%	15.87%	3.89%	1.90%
2014	21.80%	16.70%	5.98%	2.11%
2015	22.12%	17.22%	5.00%	2.40%
2016	23.33%	17.87%	5.28%	2.82%

Appendix F: Data for Figure 2

	Articles with only Chinese addresses	Articles with both Chinese and other countries' addresses	Articles with no Chinese address but at least one Chinese name	Articles with at least one Chinese address or name
2000	4.05%	0.91%	7.47%	12.43%
2001	5.07%	0.89%	7.80%	13.77%
2002	4.94%	1.05%	7.99%	13.97%
2003	5.60%	1.36%	8.12%	15.07%
2004	8.18%	1.60%	8.90%	18.67%
2005	11.66%	1.74%	8.45%	21.85%
2006	12.40%	1.87%	8.69%	22.96%
2007	12.38%	1.94%	8.64%	22.95%
2008	13.02%	2.15%	8.20%	23.37%
2009	13.51%	2.38%	7.26%	23.15%
2010	13.75%	2.62%	8.08%	24.45%
2011	13.92%	2.86%	9.02%	25.80%
2012	14.92%	3.13%	6.27%	24.32%
2013	16.08%	3.46%	7.29%	26.83%
2014	16.89%	3.82%	11.59%	32.30%
2015	17.41%	4.26%	10.51%	32.17%
2016	18.07%	4.70%	11.86%	34.64%

Appendix G: Data for Figure 5

	Scopus	Total CNKI	CNKI - Overlaps
1980	541,698	109,249	109,155
1981	549,516	138,636	138,357
1982	554,239	162,013	161,671
1983	559,936	176,620	176,328
1984	594,956	195,520	195,156
1985	613,603	225,563	225,190
1986	602,274	246,443	246,132
1987	612,467	266,789	266,374
1988	636,612	278,574	277,061
1989	654,030	286,238	284,683
1990	635,897	296,655	295,504
1991	645,502	308,295	306,826
1992	641,934	326,393	324,393
1993	673,464	341,438	336,181
1994	703,390	505,387	501,336
1995	712,437	497,619	492,527
1996	786,149	577,481	568,224
1997	792,808	602,750	590,837
1998	790,976	644,149	630,290
1999	791,630	726,419	712,836
2000	817,893	800,119	784,339
2001	855,023	834,626	811,485
2002	831,186	885,089	866,367
2003	813,112	967,012	947,388
2004	858,079	1,010,611	974,444
2005	926,921	1,112,958	1,046,518
2006	1,037,417	1,242,750	1,167,852
2007	1,122,724	1,336,371	1,260,964
2008	1,187,736	1,445,148	1,364,805
2009	1,269,105	1,529,567	1,443,529
2010	1,299,645	1,570,425	1,485,336
2011	1,385,539	1,586,152	1,504,444
2012	1,432,562	1,626,472	1,545,589
2013	1,550,456	1,683,181	1,594,313
2014	1,621,464	1,683,379	1,596,139
2015	1,647,751	1,658,766	1,573,243
2016	1,627,568	1,640,565	1,559,434