Assessment of science and technology literature of China and India as reflected in the *SCI/SSCI*

Ronald N. Kostoff, Michael E. Eriggs, Robert L. Rushenberg, Christine A. Eowles, Sujit Bhattacharya, Dustin Johnson, Alan S. Icenhour, Kimberly Nikodym, Ryan B. Barth, Simha Dodbele and Michael Pecht*

The science and technology (S&T) literature of India and China was analysed individually and then compared. Records having at least one author with an India or China address were retrieved from the Science Citation Index/Social Science Citation Index databases. Bibliometrics and computational linguistics analyses were performed on these retrieved records to generate the technical infrastructure (key authors, centres of excellence, etc.) and technical structure (pervasive technical thrusts, relationships among thrusts, etc.) of the Chinese and Indian S&T literature, and to compare them. A variety of analytical techniques were used to examine the infrastructure and technical structure from multiple perspectives, including factor analysis, correlation mapping, and document clustering (both traditional and fuzzy). The summary results and conclusions from these studies are described. The first section of the article presents an assessment of S&T literature of China. The second section presents an assessment of S&T literature of India, and the third section compares research outputs of India and China from multiple perspectives.

Keywords: Bibliometrics, research and technology assessment, science and technology, text mining.

SCIENCE and technology (S&T) are playing an increasingly important role in modern economies and defence establishments. In both cases, S&T allow: automation to be substituted for human labour; augmentation of human labour for enhanced capabilities; production of goods more rapidly and cheaply; and more complex products and processes to be produced and controlled. In order to maintain competitive advantage in both sectors, it is critical for any country to understand what other countries are producing in S&T, and what intrinsic S&T capabilities are being developed.

Disclaimer: The views presented in this article are solely those of the authors, and do not necessarily represent the views of the Department of the Navy, USA or any of its components, the University of Maryland, USA; DDL-OMNI Engineering, LLC, USA, or the National Institute of Science, Technology and Development Studies, New Delhi, India.

Ronald N. Kostoff, Dustin Johnson, Alan S. Icenhour, Kimberly Nikodym and Simha Dodbele are in the Office of Naval Research, 875 N. Randolph St., Arlington, VA 22217, USA; Dustin Johnson is presently at Northrop Grumman TASC, 12015 Lee Jackson Highway, Fairfax, VA 22033, USA; Michael B. Briggs is in the Marine Corps Warfighting Laboratory, 3255 Meyers Ave, Quantico, VA 22134, USA; Robert L. Rushenberg, Christine A. Bowles and Ryan B. Barth are in the DDL-OMNI Engineering, LLC, 8260 Greensboro Drive, Suite 600, Mclean, VA 22102, USA; Sujit Bhattacharya is in the National Institute of Science, Technology and Development Studies, Pusa Gate, K.S. Krishnan Marg, New Delhi 110 012, India; Michael Pecht is in the University of Maryland, College Park, MD 20742, USA.
*For correspondence. (e-mail: kostofr@onr.navy.mil)

This article examines the S&T literature of India and China. Both these countries had agrarian economies with large populations, facing problems of low agricultural productivity, technological backwardness, etc. However, major transformations took place in these two countries with the opening of their economies; China in the late 1980s and India in the early 1990s. Both countries are now pushing for technologically based economic growth and are demonstrating capabilities in highly complex technologies that were previously the forte of developed economies. They are striving to assert their presence in the world market[1]. We believed that it was important at this time to gauge the S&T capabilities of these two dynamic economies. In this context, we decided to assess the quantity and quality of their research output as well as examine trends in their S&T capabilities.

Table 1 provides some idea of their intrinsic resources, using USA as a basis for comparison.

Table 1. Comparison of India–China–USA resources (2006)

Resource	India	China	USA
Land mass (million sq. km)	3.29	9.6	9.63
Population (billion)	1.1	1.31	0.298
Med age (years)	24.9	32.7	36.5
Birth rate/1000	22	13.3	14.1
GDP(1) ($T)	3.61	8.86	12.36
GDP(2) ($T)	0.72	2.23	12.49

1. Purchasing power parity; 2, Official exchange rate.

Table 1 highlights some striking characteristics. India and China have large populations, but the similarities end there. China's GDP is more than double that of India, and about 70% that of USA. India's population is substantially younger than that of China, and China's population in turn is younger than that of USA. Moreover, India's birth rate is almost double that of China, or USA.

In recent years, both India and China have expanded their efforts in S&T. Figure 1 compares the number of research articles by Chinese and Indian authors over time. Data were taken from the *Science Citation Index/Social Science Citation Index* (*SCI/SSCI*)[2] and the *Engineering Compendex* (*EC*)[3].

It should be emphasized that the number of publications of a country in the *SCI/SSCI* at any point in time is a function of many variables. The number of publications is affected by research sponsorship, research productivity, quality of research, how fundamental is the research, incentives for publication, and the quality of journals in which the researchers publish. The latter is particularly important, since new higher quality journals are always being added to the *SCI/SSCI* and lower quality journals are always being removed.

Thus, if a country shows an increase in the number of *SCI/SSCI* publications over time, it could be due to some combination of increased research sponsorship, increased research productivity, higher quality research, more fundamental research, more incentives for publication, and publishing in higher quality journals that can be accessed by the *SCI/SSCI*. The same holds true for the development-oriented EC, with the exception that fundamentality of research is no longer a requirement, being replaced by quality of development.

Prior to 1980, Indian publications greatly exceeded those of the Chinese. Then, starting in 1980, Indian publications plateaued for about fifteen years. In 1980 – about the time Chinese articles began to appear in the databases – the Chinese publication rates began an exponential rise at a rate of about 20% increase per year. Indian and Chinese publications were about equal in 1995. From the period 1995–2005, the growth in Chinese publications has greatly exceeded that of India.

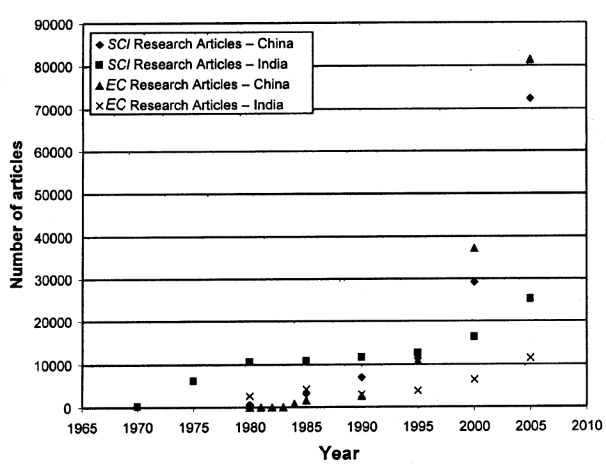

Figure 1. Number of research articles with Chinese or Indian authors.

A number of questions arise about this expanded research activity:

- What is its quality?
- What is its relevance to economic and military missions?
- What is its technical scope?
- What are the strategic areas of emphasis?
- What are the core competencies and critical masses?
- Where is it being performed?
- Who are the key performers?
- Who are the main external collaborators? What benefits result from these collaborations?
- Where is the research being published? Who is citing it?
- Do India and China display awareness of external S&T?

This article attempts to answer many of these questions using suites of tools and processes known collectively as text mining. The article is based on studies of the Chinese and Indian S&T literature published mainly in the *SCI/SSCI*, and secondarily in the *EC*. While much of the technical thrust analyses derive from papers published in the 2005 time-frame, trend analyses examined papers published starting in 1980.

Approach

Bibliometrics and computational linguistics analyses were performed on these retrieved records to generate the technical infrastructure (key authors, centres of excellence, etc.) and technical structure (pervasive thrusts, relationships among thrusts, etc.) of the Chinese and Indian S&T literature, and to compare them. A variety of specific analytical techniques were used to examine the infrastructure and technical structure from multiple perspectives, including factor analysis, correlation mapping, and document clustering (both traditional and fuzzy). The summary results and conclusions from these studies follow. The first section presents an assessment of S&T literature of China. The second section presents an assessment of S&T literature of India, and the third section compares research outputs of India and China from multiple perspectives. For more detailed presentation of the methodologies and results from the total India and China studies, see Kostoff *et al.*[4–9].

Results

China

- China's publication of *SCI/SSCI* research articles had an annual exponential growth rate of 20% over the last 25 years.

- The major themes of these articles have shifted gradually over time from multidisciplinary science, medicine, and life sciences in 1980 to materials, chemistry, and physics in 2005, in that order.
- The major journals in which Chinese researchers publish have Impact Factors (IFs; a proxy metric for quality) that average almost an order of magnitude less than those of the major journals they reference. These major publication journals have a high domestic component and, on average, relatively recent *SCI*/*SSCI* access dates. Beyond increased productivity and recent sponsorship, recent *SCI*/*SSCI* access by numerous journals must be considered as a contributor to the apparent high *SCI*/*SSCI* growth rate of Chinese research articles.
- However, while the present median IF of the major journals in which Chinese researchers publish is relatively low, the time trend is positive. The presence of Chinese-authored papers in very high quality journals, on average, is outpacing overall Chinese research article growth.
- Collaboration with external countries produces a substantial increase in the number of Chinese articles published in very high IF journals, but produces negligible change in median citations of overall Chinese publications. Major impact of the collaboration is at the high end.
- The major Chinese research article contributor, the Chinese Academy of Science, has common thematic interests with many of the Chinese institutions, but does not co-publish to the same extent.
- Many of the institutions not connected to the main Chinese network through common thematic interests are from Hong Kong, suggesting the remnants of some degree of technical thematic independence of Hong Kong from the more traditional Mainland institutions.
- When comparing China–Japan and China–USA collaborations, the collaborative research relationship with Japan appears to be more quid pro quo than the one with USA, given the emphasis of Japanese articles on nanotechnology first followed by biomedical, and emphasis of USA articles on biomedical first followed by nanotechnology.
- In comparing highly cited to poorly cited Chinese-authored documents, it was found that:

(i) Highly cited documents are published in international journals, whereas poorly cited documents are published in domestic Chinese journals. The broad technical areas in the highly and poorly cited groups are relatively similar.

(ii) Historically, most of the author names associated with the highly cited documents were non-Asian, suggesting that the Chinese participants were contributors to larger non-Chinese-driven research programmes. In recent years, Chinese participation in the most cited papers has increased to about half. The names associated with the least cited papers are all Chinese.

(iii) In recent years, institutions associated with the most cited papers are roughly half Chinese/half USA, showing the disproportionate contribution of the US institutions to highly cited Chinese research. Further, the two main Hong Kong universities have an order of magnitude higher relative representation (based on total publications) on the highly cited papers than their most visible Mainland counterparts (Chinese Academy of Science, Tsing Hua University, Zhejiang University).

(iv) USA is the predominant collaborator for the highly cited papers, having about five times the relative representation expected based on its collaboration with all Chinese research.

- USA is a key collaborator with China in almost every technical category identified in the technical structure taxonomy, with Japan being a close second. Malaysia was the strongest collaborator in the 'Crystal and Chemical Structure Synthesis' category.
- The difference in relative thematic emphasis between USA and China is dramatic! Relative to USA, China emphasizes the physical and engineering sciences that underpin defence and commercial needs. Relative to China, USA emphasizes research areas focused on medical, psychological and social problems. There are even focused research areas where China leads USA in absolute number of research articles published. In such areas, China's investment allocation relative to that of USA is greater than four. Moreover, when the time trends in these areas are examined, the much higher Chinese gradients relative to those of USA imply that the *gap will only widen in the future*.
- China has expanded its documented research output dramatically in the last decade. Its citation performance, based on the present country assessment and other specific technology assessments, is competitive with that of other developing nations, but not yet competitive with that of the developed nations. It is not clear whether this non-competitiveness is due to overly applied research, lower quality research, both or neither. Also, since there is a lag period associated with citation analysis, these conclusions apply to vintage research at least five years old. Whether they apply to very recent research will only be known at the end of the decade. To resolve this non-competitiveness issue, experts are required to sample similar articles written by Chinese and non-Chinese authors in a number of disciplines, compare these article pairs for quality and level of development, and correlate them with citations. While resource-intensive, this next step is required to resolve the quality/citation issue.

India

- India's research article contribution reached a plateau during the period 1980–1995, and it has since started a rapid increase.

• The main technical focus at present is on the three physical sciences areas of chemistry, physics and materials, supported by a strong foundation of applied mathematics.

• Half the journals that contain the most Indian papers are domestic Indian journals, and they have low IF.

• Collaboration with external researchers has the effect of dramatically increasing the presence of papers with Indian authors in the higher IFs journals, as well as the number of papers with high citations.

• India is increasing its growth of articles in highly cited journals at a faster rate than its overall increase in growth of research articles.

• Journals most cited by Indian authors are all international, and have IFs an order of magnitude larger than those of journals that publish the most Indian papers.

• The network of Indian co-publishing institutions is weakly linked, but the network of institutions with common thematic interests has some strong links. In particular, the Indian Institutes of Technology (IITs) and the Indian Institute of Science (IISc; India's two leading research publishing institutions) have no strong co-publishing links, but they are at the centres of a bi-polar core of the network of institutions with common thematic interests. Two CSIR laboratories, namely Indian Institute of Chemical Technology and National Chemical Laboratory, also exhibit thematic links with the above network.

• In comparing the most cited Indian papers with the least cited, the following characteristics were identified:

(i) The most cited were published in international journals, while the least cited are published in domestic Indian journals. The most cited emphasized chemistry, physics and medicine, while the least cited had substantial representation from agricultural and veterinary sciences.

(ii) The authors of the least cited papers are all Indian. The authors of the most cited (published a few decades ago) are mainly Indian, but due to an anomaly, the names of the top authors of highly cited papers published relatively recently are non-Indian.

(iii) Institutions associated with the least cited papers are mainly Indian, and this has not changed with time. For papers published decades ago, institutions associated with the most cited papers are mainly Indian, and are dominated by the IITs and IISc. For relatively recent papers, institutions associated with the most cited papers are non-Indian, due to the same anomaly referenced above.

(iv) USA has been the leading collaborator on the most highly cited papers for decades, increasing its participation from 28% in 1979–87 to 65% in 1998–2003. India has remained essentially the only country associated with the least cited papers, with a handful of countries listed with very low frequency in recently published papers.

• In comparing the research investment allocation of India relative to that of USA, India's strong relative emphasis is on traditional agricultural products, phenomena in the visible spectrum, and selected chemistry topics. The relative research emphasis areas of USA focused on medical, psychological and sociological.

China–India comparison

A comparison was made of the research output literature for India and China. The following conclusions are drawn.

• China has increased its total research article output by two orders of magnitude since 1980, whereas India has increased its research article output by 2.5 over the same time period, a factor of forty difference (Figure 1).

• Both countries have strong focus on applied mathematics and nanotechnology.

• For both countries, majority of journals containing their most publications are domestic low IF journals. The median IFs of the journals containing the largest number of USA publications are almost an order of magnitude larger than the similar journals for China or India.

• The median initial *SCI* access date of the journals containing the largest number of India publications is 1973, whereas the analogous date for Chinese publications is 1995. Thus, at least some of the excess growth of China's research publications relative to that of India must have come from additional journals being accessed by the *SCI*/*SSCI*, rather than purely increased productivity or increased research sponsorship.

• Since 1985, the weighted IF (based on top ten publishing journals) for Chinese articles has been greater than that for India. The Chinese are not only publishing substantially more than India, but they are publishing in higher IF journals as well.

• For the illustrative example of three key journals in physics, chemistry and biology, publications by Chinese authors have significantly outpaced those by Indian authors since 2000 (Figure 2).

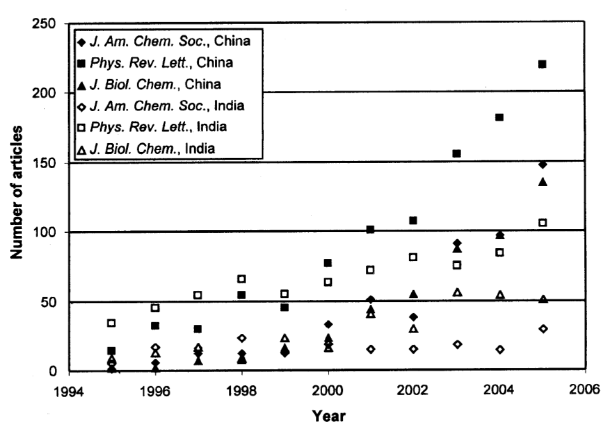

Figure 2. Number of publications by Chinese or Indian authors in selected high impact factor journals as a function of time.

- Both China and India benefit dramatically from external research collaboration by drastically increasing their publications in the highest IF journals.
- The main collaborators are the leading technology countries, especially USA, Japan, Germany and England. For India, USA increased its participation from less than 2% of papers in 1980 to 7% in 2005, and for China, USA increased its participation from 5% in 1980 to 8% in 2005.
- China outperforms India at the high citation end, independent of whether the metric used is papers with more than 100 citations, median of top twenty papers, or median of top 1% of papers.
- In comparing the most cited papers for each country, the following was found:

(i) The most cited papers were published in international journals, while the least cited were published mainly in domestic journals of the respective countries.

(ii) Veterinary and agricultural sciences tended to be over-represented in India's poorly cited research.

(iii) Names associated with the least cited papers for India and China were all Indian and Chinese respectively, while those associated with the highly cited papers were mixed.

(iv) Institutions associated with the least cited papers for India and China were exclusively domestic.

(v) The two main Hong Kong universities had an order of magnitude higher relative representation (based on total publications) on the highly cited papers than their most visible Mainland counterparts.

- China's technical emphasis is on materials (especially nanostructures), applied mathematics (including nonlinear dynamics), and luminescent phenomena. Biomedicine has less emphasis, except for local diseases.
- India places more relative effort on biomedicine, slightly less than China in applied mathematics, has strong interest in thin films and nanotechnology, physics strongly associated with surfaces/materials and optical spectra phenomena, and greater relative emphasis on veterinary and agricultural sciences.
- China outperforms India in citations across all major technical categories. In turn, Australia outperforms China in citations across the same categories using the same types of metrics.
- Relative to India, China emphasizes research in nanotechnology, mathematics (especially nonlinear dynamics), cancer and viral disease treatment, and luminescent phenomena.
- Relative to China, India emphasizes on agriculture, veterinary science, tropical diseases, nuclear power and optical phenomena.
- The bottom line in this comparison is unmistakeable. In 1980, India was light years ahead of China in volume and breadth of published research. For almost two decades, India's research output production stagnated. During that period, China's research production increased exponentially. Presently, China outperforms India substantially both in quantity and quality (as measured by the IF and relative citations of research output). The gap is widening and shows no sign of abating, if present research policies are continued!

1. Bhattacharya. S, and Nath, P., Using patent statistics as a measure of 'technological assertiveness': A China–India comparison. *Curr. Sci.*, 2002, **83**, 23–29.
2. *SCI*, Certain data included herein are derived from the Science Citation Index/Social Science Citation Index prepared by the Thomson Scientific®, Inc. (Thomson®), Philadelphia. Pennsylvania, USA. 2006; © Copyright Thornson Scientific® 2006: All rights reserved.
3. *EC*, Certain data included herein are derived from the Engineering Compendex, accessed through Engineering Village, which is a product of Elsevier Engineering Information, Inc, Haboken, NI, a wholly owned subsidiary of Elsevier Science, 2006.
4. Kostoff, R. N., Briggs, M., Rushenberg, R., Bowles, C. and Pecht, M., The structure and infrastructure of Chinese science and technology. DTIC Technical Report Number ADA443315 (http://www.dtic.mil/), Defense Technical Information Center, VA, USA, 2006.
5. Kostoff, R. N., Johnson, D., Bowles. C. A. and Dodbele. S., Assessment of India's research literature. DTIC Technical Report Number ADA444625 (http://www.dtic.mil/), Defense Technical Information Center, VA, USA, 2006.
6. Kostoff, R. N., Bhattacharya. S. and Pecht, M., Assessment of China's and India's science and technology literature – introduction, background, and approach. *Technol. Forecast. Social Change*, 2007; 1016/j.techfore.2007.02.004 (in press).
7. Kostoff, R. N. *et al.*, Assessment of India's research literature. *Technol. Forecast. Social Change*, 2007; 1016/j.techfore.2007.02.009 (in press).
8. Kostoff, R. N. *et al.*, Chinese science and technology – Structure and infrastructure. *Technol. Forecast. Social Change*, 2007; 1016/j.techfore.2007.02.008 (in press).
9. Kostoff, R. N. *et al.*, Comparisons of the structure and infrastructure of Chinese and Indian science and technology. *Technol. Forecast. Social Change*, 2007; 1016/j.techfore.2007.02.007 (in press).

Received 22 February 2007; revised accepted 16 July 2007