Available online at www.sciencedirect.com
ScienceDirect
Technological Forecasting & Social Change 74 (2007) 1539–1573
Technological Forecasting and Social Change

Chinese science and technology — Structure and infrastructure

Ronald N. Kostoff [a,*], Michael B. Briggs [b], Robert L. Rushenberg [c], Christine A. Bowles [c], Alan S. Icenhour [a], Kimberley F. Nikodym [a], Ryan B. Barth [c], Michael Pecht [d]

[a] *Office of Naval Research, 875 N. Randolph St., Arlington, VA 22217, USA*
[b] *Marine Corps Warfighting Laboratory, 3255 Meyers Ave, Quantico, VA 22134, USA*
[c] *DDL-OMNI Engineering, LLC, 8260 Greensboro Drive, Mclean, VA 22102, USA*
[d] *University of Maryland, College Park, MD 20742, USA*

Received 28 October 2006; accepted 16 February 2007

Abstract

This paper identifies and analyzes the science and technology core competencies of China, based on a sampling of approximately half of the total Chinese publication output in the Science Citation Index/ Social Science Citation Index (SCI/SSCI) [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: © Copyright Thomson Scientific® 2006. All rights reserved. [1]] for 2005. Aggregate China publication and citation bibliometrics were obtained and a hierarchical research taxonomy, based on document clustering, was generated. Additionally, bibliometrics and thematic trends were tracked over the past two decades.

The key findings were that China's output of research articles has significantly expanded in the last decade. In terms of sheer numbers of research articles, especially in cuting-edge technologies, such as nanotechnology and energetic materials, it is among the leaders. Compared to the USA, the bulk of China's articles focus on the physical and engineering sciences, while the USA articles (compared to China) focus on medical, social, and psychological sciences.
Published by Elsevier Inc.

Keywords: China; Science and technology; Bibliometrics; Citation analysis; Impact Factor; Computational linguistics; Core competencies; Research evaluation; CLUTO; Nanotechnology; Clustering; Taxonomies

* Corresponding author. Tel.: +1 703 696 4198; fax: +1 703 696 8744.
E-mail address: kostofr@onr.navy.mil (R.N. Kostoff).

0040-1625/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.techfore.2007.02.008

1. Background

There are many reasons to understand the state and trends in a county's science and technology (S&T). For example, Kostoff et al. [2] conducted a text mining assessment of the global nanotechnology literature including that of the USA, India and China. Pecht [3] noted that China has made significant achievements in hi-tech areas that are keys to economic development and has begun to gain ground in the world's S&T community.

This paper presents the text mining of a sub-set of China's national research core competencies in terms of the identification of China's main research thrusts, volume of research output in main research thrusts, and bibliometric and thematic temporal trends.

2. Approach and results

In this paper, two types of results are presented, bibliometrics and taxonomies. Bibliometrics provide an indication of the technical infrastructure (prolific authors, journals, institutions, citations), while taxonomies provide an indication of major technology thrusts and their relationships.

Fig. 1 shows the number of SCI and Engineering Compendex (EC) [4] research articles per year as a function of time. The SCI data is denoted by diamonds, while the EC data is denoted by squares. A best fit curve to the SCI data (see Fig. 1) shows an exponential annual growth rate of approximately 20% over the 25 year time period.

The SCI output contains a field called Subject Category. It is essentially a classification by technical thrust of each article. Table 1 lists the top ten Subject Categories for all of the Chinese research articles since 1980. There has been a gradual shift of emphasis from multidisciplinary science, medicine, and life science in 1980 to that of materials, chemistry, and physics respectively in 2005. This correlates with China's emphases on the physical and engineering sciences relative to the USA's emphases on medical, sociological, and psychological sciences, a point that will be discussed in the last section of this paper.

The EC contains fields called Classification Code and Controlled Vocabulary, both of which are similar to the SCI Subject Category field. Table 2 lists the top ten classification codes for all of the Chinese

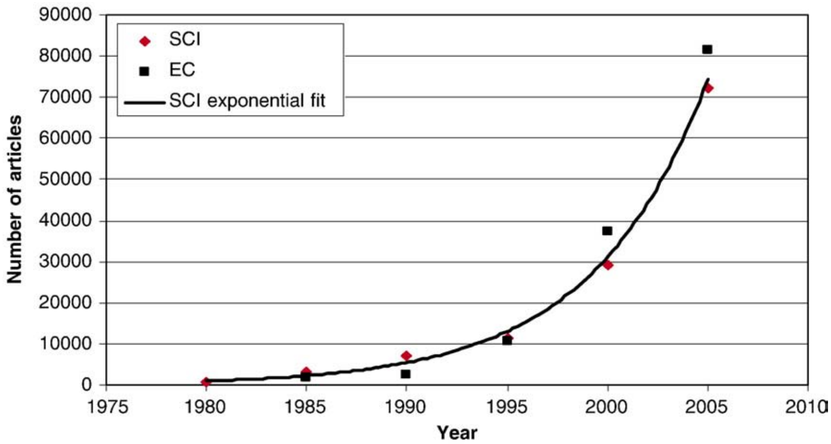

Fig. 1. Number of SCI and EC research articles annually with at least one Chinese author.

Table 1
SCI subject categories as a function of time

1980		1985		1990		1995		2000		2005	
Multidisciplinary sciences	130	Chemistry, multidisciplinary	348	Chemistry, multidisciplinary	652	Chemistry, multidisciplinary	1354	Materials Science, multidisciplinary	2712	Materials Science, multidisciplinary	7091
Medicine, general and internal	123	Physics, multidisciplinary	279	Physics, multidisciplinary	621	Materials Science, multidisciplinary	976	Chemistry, multidisciplinary	2285	Chemistry, physical	4653
Zoology	51	Multidisciplinary sciences	263	Physics, condensed matter	485	Physics, multidisciplinary	863	Physics, multidisciplinary	1792	Physics, multidisciplinary	4478
Entomology	49	Pharmacology and pharmacy	187	Physics, applied	471	Physics, applied	790	Physics, applied	1655	Chemistry, multidisciplinary	4301
Astronomy and Astrophysics	43	Mathematics	186	Mathematics	454	Physics, condensed matter	762	Chemistry, physical	1546	Physics, applied	3823
Paleontology	35	Medicine, general and internal	149	Multidisciplinary sciences	453	Mathematics	675	Physics, condensed matter	1287	Computer Science, theory and methods	3348
Geology	30	Physics, applied	105	Materials Science, multidisciplinary	448	Chemistry, physical	616	Chemistry, analytical	1248	Metallurgy and Metallurgical Engineering	3093
Physics, multidisciplinary	24	Geosciences, multidisciplinary	104	Mathematics, applied	356	Mathematics, applied	528	Mathematics, applied	1197	Biochemistry and Molecular Biology	2789
Geochemistry and Geophysics	23	Engineering, electrical and electronic	99	Medicine, general and internal	349	Chemistry, analytical	461	Engineering, electrical and electronic	1181	Physics, condensed matter	2738
Physics, applied	14	Geochemistry and geophysics	98	Pharmacology and Pharmacy	256	Multidisciplinary sciences	448	Biochemistry and Molecular Biology	1152	Crystallography	2585

Table 2
Engineering compendex classification code for Chinese journal articles as a function of time

1985		1990		1995		2000		2005	
Applied Mathematics	388	Applied Mathematics	551	Numerical Methods	1042	Inorganic Compounds	2119	Chemical Reactions	9188
Chemical Products Generally	360	Applied Physics Generally	536	Inorganic Compounds	769	Applied Mathematics	2094	Computer Applications	8639
Applied Physics Generally	343	Chemical Products Generally	532	Chemical Reactions	747	Physical Properties of Gases, Liquids and Solids	2043	Applied Mathematics	8228
Chemical Apparatus and Plants; Unit Operations; Unit Processes	241	Strength of Building Materials; Mechanical Properties	359	Physical Properties of Gases, Liquids and Solids	690	Numerical Methods	1976	Numerical Methods	7927
Metallurgy and Metallography	216	Metallurgy and Metallography	324	Light/Optics	647	Chemical Reactions	1907	Inorganic Compounds	7921
Strength of Building Materials; Mechanical Properties	210	Electricity and Magnetism	310	Strength of Building Materials; Mechanical Properties	637	Computer Applications	1716	Physical Properties of Gases, Liquids and Solids	7116
Light, Optics and Optical Devices	207	Light, Optics and Optical Devices	285	Optical Devices & Systems	625	Chemical Operations	1447	Chemical Operations	6974
Chemistry	179	Computer Software, Data Handling and Applications	250	Physical Chemistry	564	Mechanics	1403	Light/Optics	6702
Iron and Steel	169	Chemistry	250	Mechanics	515	Light/Optics	1363	Organic Compounds	6114
Electricity and Magnetism	136	Chemical Apparatus and Plants; Unit Operations; Unit Processes	221	Applied Mathematics	481	Organic Compounds	1254	Strength of Building Materials; Mechanical Properties	5671

Table 3
Engineering compendex controlled vocabulary for Chinese journal articles as a function of time

1985		1990		1995		2000		2005	
Mathematical Models	92	Mathematical Models	100	Mathematical Models	731	Mathematical Models	1959	Computer Simulation	6873
Mathematical Techniques — Finite Element Method	31	Fracture Mechanics	65	Computer Simulation	337	Computer Simulation	1521	Mathematical Models	6632
Microscopic Examination	27	Computer Simulation	46	Thermal Effects	318	Algorithms	817	Algorithms	3224
Microscopes, Electron-Applications	27	Ceramic Materials	39	Calculations	288	Synthesis (Chemical)	711	Synthesis (Chemical)	3095
Fracture Mechanics	22	Stresses — Analysis	37	Synthesis (Chemical)	247	Thermal Effects	598	Nanostructured Materials	2806
Computer Simulation	21	Stresses	36	Algorithms	226	Optimization	507	Scanning Electron Microscopy	2592
Chemical Reactions — Reaction Kinetics	21	Mathematical Techniques	34	Microstructure	202	X Ray Diffraction Analysis	467	X Ray Diffraction Analysis	2507
Thermodynamics	19	Elasticity	34	Numerical Analysis	201	Finite Element Method	464	Optimization	2463
Stresses — Analysis	18	Computer Programming — Algorithms	33	Laser Applications	190	Nanostructured Materials	436	Finite Element Method	1919
Heat Treatment — Annealing	18	Mathematical Techniques — Finite Element Method	33	Composition	188	Scanning Electron Microscopy	425	Transmission Electron Microscopy	1831

research articles since 1985. Table 3 lists the Controlled Vocabulary for the same time period. Table 2 shows a shift towards chemistry and numerical modeling, while many of the topics listed for 2005 in Table 3 indicate an emphasis on computation and nanotechnology. One constant over time has been the relatively strong emphasis on computer simulation and mathematical modeling.

The twenty journals containing the most papers with at least one China author during the period 2004–2005 are listed in Table 4. The first column is the full journal name, the second column is the number of papers in the journal from the database, the third column is the journal's Impact Factor (the Impact Factor is the ratio of cites of recent articles to numbers of recent articles, and can be considered one measure of a journal's ability to attract citations), the fourth column is the journal's theme, and the fifth column provides the initial SCI access date for the journal. These journals appear to be concentrated in chemistry, materials, and physics, with one journal about medicine. Many of the articles are published in domestic Chinese journals and the initial SCI access dates are relatively recent (median initial access date 1995)—both facts are consistent with the lower Impact Factors for many of the journals.

The variation with time in the weighted Impact Factor for journals containing Chinese articles is shown in Fig. 2. The weighted Impact Factor is calculated by evaluating the Impact Factor for the top ten journals in a

Table 4
Most prolific Chinese journals — 2004–2005

China journal	# Papers	Impact Factors	Theme	SCI Access Date
Acta Crystallographica Section E—Structure Reports Online	1494	0.581	Materials	2001
Acta Physica Sinica	1032	1.051	Physics	1999
Chinese Physics Letters	920	1.276	Physics	1989
Rare Metal Materials And Engineering	872	0.400	Materials	1997
Spectroscopy And Spectral Analysis	610	0.557	Physics	1999
Physical Review B	547	3.185	Physics	1964
Chemical Journal Of Chinese Universities—Chinese	531	0.771	Chemistry	1995
Materials Letters	528	1.299	Materials	1985
Priem 5: The Fifth Pacific Rim International Conference On Advanced Materials And Processing, Pts 1–5	520	0.000	Materials	2005
Chinese Science Bulletin	513	0.783	Science	1989
Applied Physics Letters	509	4.127	Physics	1962
Chinese Chemical Letters	504	0.355	Chemistry	1995
Chinese Physics	497	1.256	Physics	1981
Transactions of Nonferrous Metals Society of China	479	0.302	Materials	1995
Chinese Journal of Analytical Chemistry	459	0.397	Chemistry	1999
Communications in Theoretical Physics	453	0.872	Physics	1985
Acta Chimica Sinica	423	0.845	Chemistry	1980
Chinese Journal Of Inorganic Chemistry	418	0.697	Chemistry	1999
Chinese Medical Journal	410	0.561	Medical	1964
Journal of Physical Chemistry B	405	4.033	Chemistry	1997
High-Performance Ceramics iii, Pts 1 and 2	398	0.000	Materials	2005
Journal of Rare Earths	346	0.249	Materials	1995
Physics Letters A	340	1.550	Physics	1979
Journal of Crystal Growth	338	1.681	Materials	1971
Chinese Journal Of Chemistry	323	0.819	Chemistry	1995

Impact Factors based on 2004 data.

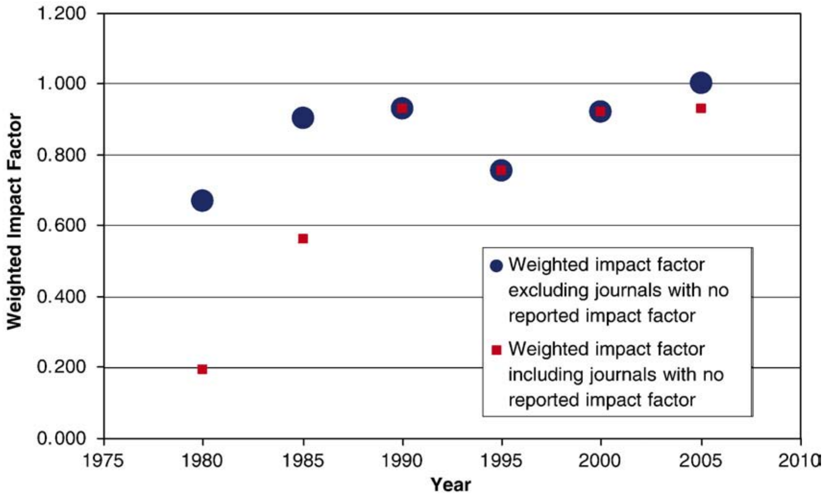

Fig. 2. Weighted Impact Factor for the top ten journals with Chinese authors.

given year and then calculating an average that is weighted by the number of publications in each journal. The circles reflect the computations of weighted Impact Factors that *omit* journals with no reported Impact Factor, while the squares *include* the journals with no reported Impact Factor (i.e., the Impact Factor for those journals is taken as zero). These two results differ primarily in the earlier years (1980 and 1985), due to many of the publications being in more obscure journals. The minor difference in Impact Factors shown in 2005 is because of a large number of publications in a conference's proceedings, which does not have a reported Impact Factor. The overall trend seen in Fig. 2 is an increase in Impact Factor.

Fig. 3 shows the number of Chinese publications for three of the key discipline-oriented high-Impact Factor journals (Journal of the American Chemical Society, Physical Review Letters, and Journal of Biological Chemistry) as a function of time. This figure depicts a large increase in Chinese publications in higher Impact

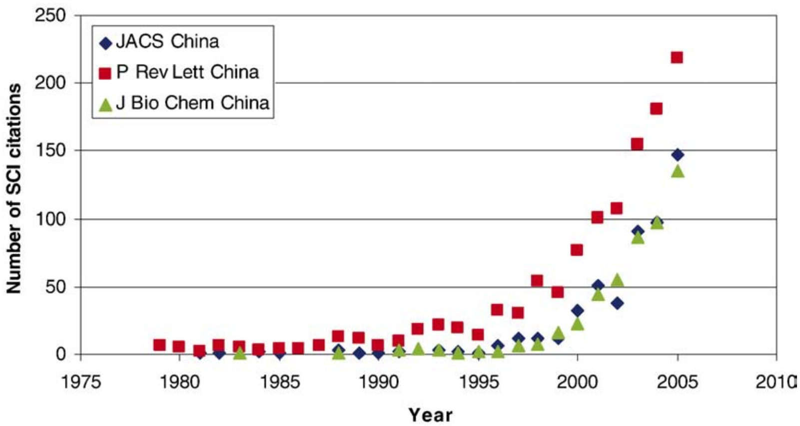

Fig. 3. Number of publications by Chinese authors in selected high impact factor journals as a function of time.

Table 5
Numbers of Chinese publications during 2005 with and without collaborators for high impact factor journals

Journal	China only	China and collaborators	Impact Factor [a]
Nature	3	19	29.273
Science	4	26	30.927
Physical Review Letters	46	219	7.489
PNAS-USA [b]	15	58	10.231

[a] Based of 2005 Impact Factors.
[b] Proceedings of the National Academy of Science.

Factor (quality) journals starting in the mid-1990s—in parallel with the overall increase in publications (Fig. 1). This publication trend is reflected in the overall increase in weighted Impact Factor shown in Fig. 2.

The numbers of publications in selected high Impact Factor journals with only Chinese authors and those for Chinese plus collaborative authors are shown in Table 5. Collaboration with authors from other countries strongly influences the numbers of publications by Chinese authors in high Impact Factor journals. The numbers of publications are increased by factors of four to six for the examples shown.

2.1. Prolific institutions

Table 6 lists the top 25 publishing institutions as a function of time: 1985, 1995, and 2005. The table shows a dramatic increase in the number of publications by many of the institutions. Throughout the period examined, the Chinese Academy of Science remained at or near the top of the list. There has been a notable shift by a number of universities towards the top of the list. For example, Tsing Hua University increased its number of publications by an order of magnitude from 1995 to 2005. The Zhejiang University and Shanghai Jiao Tong University increased publications by factors of 17 and 24, respectively, during the same period. Peking University, which is in the top 5 as of 2005, was not even in the top 25 in 1995.

The increase in numbers of publications was examined in terms of the Impact Factor. Weighted Impact Factors were calculated for the top ten institutions for 1985, 1995, and 2005 as shown in Table 7. (The weighted Impact Factor was computed based on the top ten publishing journals for an institution for a given year.) The overall trend of increasing Impact Factor shown in Fig. 2 can be seen for many of the institutions. However, there are some interesting insights from the data presented in Table 7.

A higher number of publications does not necessarily equate to higher Impact Factor journals. In fact, the institutions with the highest Impact Factors in 2005 are near the bottom of the top 10 (in terms of numbers of publications). Some of the institutions have increased both the numbers of publications and the Impact Factor (e.g., Chinese Academy of Science). However, it appears that some institutions have increased numbers of publications while sacrificing journal Impact Factor. For example, Tsing Hua University increased publications by a factor of ten from 1995 to 2005, but the Impact Factor decreased by nearly a factor of two. The Impact Factors for Zhejiang University and Shanghai Jiao Tong University for 1995 were computed to be 0.723 and 1.251, respectively. Both of these institutions were outside of the top ten in publications in 1995, but were in the top five during 2005. The Impact Factor remained about the same for Zhejiang University, while it was cut in half for Shanghai Jiao Tong University.

To help determine the co-publishing characteristics of institutions, an auto-correlation map of the thirty most prolific Chinese institutions for 2005 was generated using the TechOasis software [5]. No strongly

Table 6
Top 25 Chinese publishing institutions based on total numbers of publications for 1985, 1995, and 2005

1985		1995		2005	
Institution name	Record count (3105)	Institution name	Record count (11,397)	Institution name	Record count (72,362)
Chinese Acad Sci	870	Acad Sinica	1808	Chinese Acad Sci	14051
Acad Sinica	175	Chinese Acad Sci	1524	Tsing Hua Univ	3650
Beijing Univ	126	Nanjing Univ	617	Zhejiang Univ	3268
Chinese Acad Med Sci	99	Beijing Univ	488	Peking Univ	2710
Univ Sci & Technol China	83	Univ Sci & Technol China	358	Shanghai Jiao Tong Univ	2435
Nanjing Univ	72	Fudan Univ	353	Univ Hong Kong	2109
Fudan Univ	66	Tsing Hua Univ	346	Nanjing Univ	2031
Beijing Med Coll	42	Jilin Univ	259	Univ Sci & Technol China	1992
Nankai Univ	40	Nankai Univ	243	Fudan Univ	1770
State Seismol Bur	34	Lanzhou Univ	239	Chinese Univ Hong Kong	1743
Chinese Acad Geol Sci	29	CCAST	195	Hong Kong Polytech Univ	1411
Jilin Univ	29	Zhejiang Univ	188	Shandong Univ	1344
Lanzhou Univ	28	Wuhan Univ	183	City Univ Hong Kong	1332
Wuhan Univ	28	China Ctr Adv Sci & Technol	175	Jilin Univ	1330
Tsinghua Univ	27	Xian Jiaotong Univ	161	Harbin Inst Technol	1221
Beijing Normal Univ	26	Shandong Univ	158	Huazhong Univ Sci & Technol	1206
E China Normal Univ	26	Beijing Normal Univ	151	Wuhan Univ	1181
Shanghai First Med Coll	26	Xiamen Univ	132	Nankai Univ	1179
Zhongshan Univ	22	Zhongshan Univ	129	Tianjin Univ	1134
Beijing Univ Iron & Steel Technol	21	Huazhong Univ Sci & Technol	117	Hong Kong Univ Sci & Technol	1128
China Natl Ctr Prevent Med	21	Harbin Inst Technol	111	Xian Jiaotong Univ	1074
Huazhong Univ Sci & Technol	18	Beijing Med Univ	107	Sichuan Univ	1051
Peking Univ	18	Shanghai Jiao Tong Univ	103	Dalian Univ Technol	958
Natl Bur Oceanog	17	Shanghai Med Univ	93	Lanzhou Univ	842
Zhejiang Univ	17	Sichuan Univ	93	Beijing Normal Univ	693

connected publishing groupings are evident, but five moderately-connected publishing groupings can be identified.

- Harbin Institute of Technology-centered group (bottom center).
- Tsing Hua University and University of Science and Technology Beijing group (bottom left).
- Nankai University-centered group (top center).

Table 7
Weighted Impact Factors for top ten Chinese research institutions for selected years (order of institutions is based on the total number of publications during a given year)

1985		1995		2005	
Institution	Weighted Impact Factor	Institution	Weighted Impact Factor	Institution	Weighted Impact Factor
Chinese Acad Sci	0.529	Acad Sinica	1.267	Chinese Acad Sci	1.849
Acad Sinica	0.532	Chinese Acad Sci	1.163	Tsing Hua Univ	0.873
Beijing Univ	0.706	Nanjing Univ	1.673	Zhejiang Univ	0.741
Chinese Acad Med Sci	2.122	Beijing Univ	0.916	Peking Univ	2.626
Univ Sci & Technol China	0.566	Univ Sci and Technol China	2.092	Shanghai Jiao Tong Univ	0.667
Nanjing Univ	0.772	Fudan Univ	1.586	Univ Hong Kong	2.609
Fudan Univ	0.765	Tsing Hua Univ	1.470	Nanjing Univ	1.668
Beijing Med Coll	0.780	Jilin Univ	0.660	Univ Sci & Technol China	3.601
Nankai Univ	0.442	Nankai Univ	0.863	Fudan Univ	1.856
State Seismol Bur	0.878	Lanzhou Univ	0.844	Chinese Univ Hong Kong	3.494

- University of Hong Kong-centered group (top left).
- Chinese Academy of Science group.

A factor analysis of the top thirty Chinese institutions for 2005 was performed to assess the publishing linkages among institutions. A five factor model was selected based on the groupings shown in Fig. 4. Four groupings were identified:

- Wuhan University strongly linked with Shandong University, Sichuan University, and the University of Science and Technology China. Weaker linkages are with Shanghai Jiao Tong University, Nankai University, Zhejiang University, and Peking University.
- Chinese University of Hong Kong strongly linked to University of Hong Kong and Hong Kong Polytech University, and weakly linked with City University of Hong Kong, Hong Kong University of Science and Technology, and Sun Yat Sen University.
- Tsing Hua University weakly linked to University of Science and Technology Beijing, Harbin Institute of Technology, and Jilin University.
- Tongji University strongly linked with Shanghai University and Fudan University.

To display the thematic linkages among institutions, two institution-phrase cross-correlation maps were generated that show institutional relationships based on use of common terminology (Figs. 5 and 6). The first map, Fig. 5, includes very generic technical phrases. Because of their universal use in many papers, these generic pharases tend to blur the differentiation among papers from a grouping/clustering perspective. The second map, Fig. 6, does not include these general terms as stand-alone phrases (they may be included as part of a larger phrase), but replaces them with more detailed lower frequency phrases.

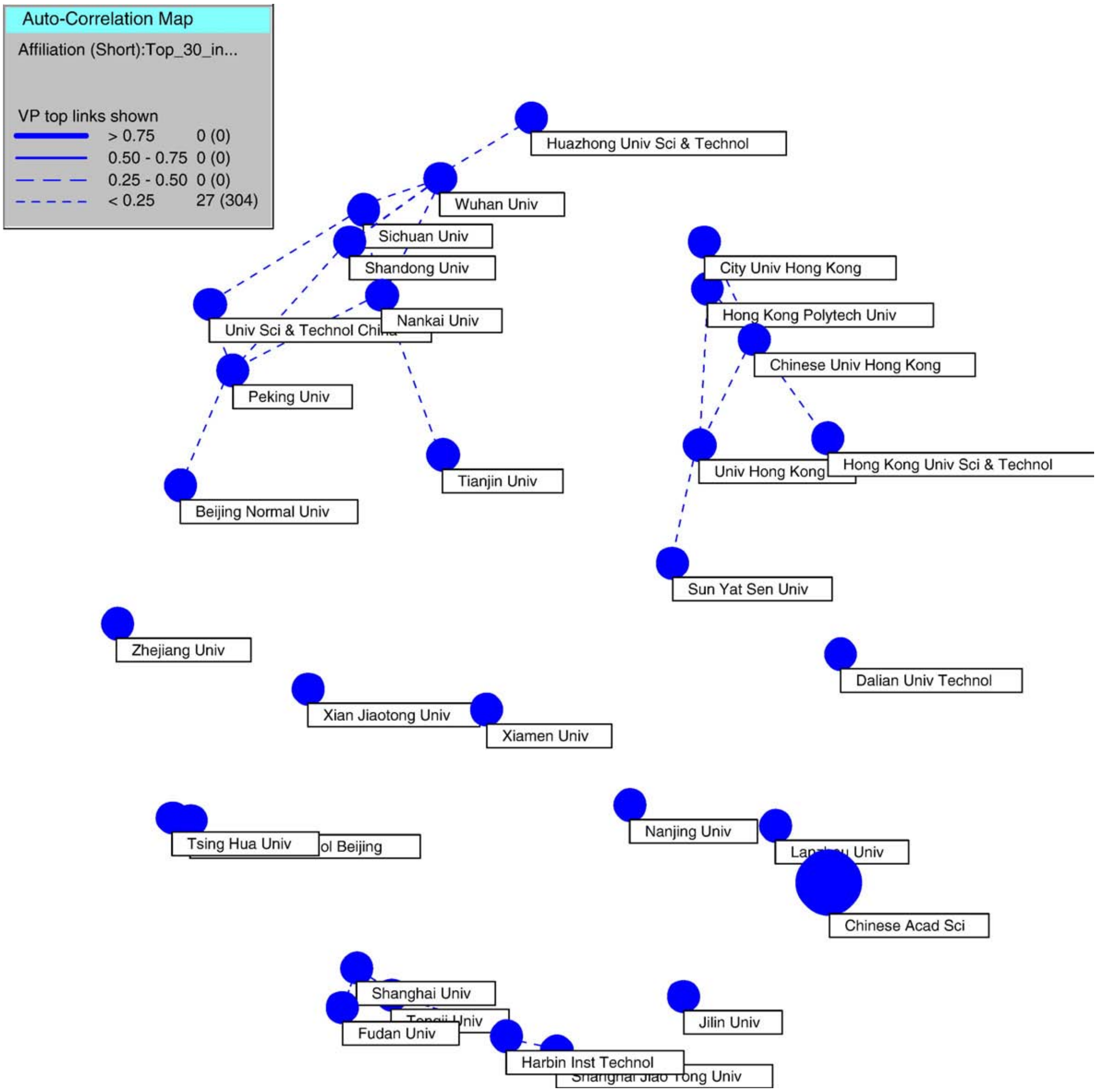

Fig. 4. Chinese top thirty publishing institutions (2005) auto-correlation map.

One immediately observable difference between the institution auto-correlation map and the institution-phrase cross-correlation maps is the number of displayed linkages and the strength of the linkages. Because of the binding effect of the generic phrases, Fig. 5 has some internal linkages that are stronger than those in Fig. 6.

There are two major differences between the auto-correlation map of Fig. 4 and the focused phrase cross-correlation map of Fig. 6. First, in Fig. 4, the Chinese Academy of Science appears isolated, while in Fig. 6 it is the central core of the institutional research network. This difference suggests that the Chinese Academy of Science has common thematic interests with many of the institutions, but does not co-publish to the same extent. Second, many of the institutions not connected to the main network of common thematic interests in

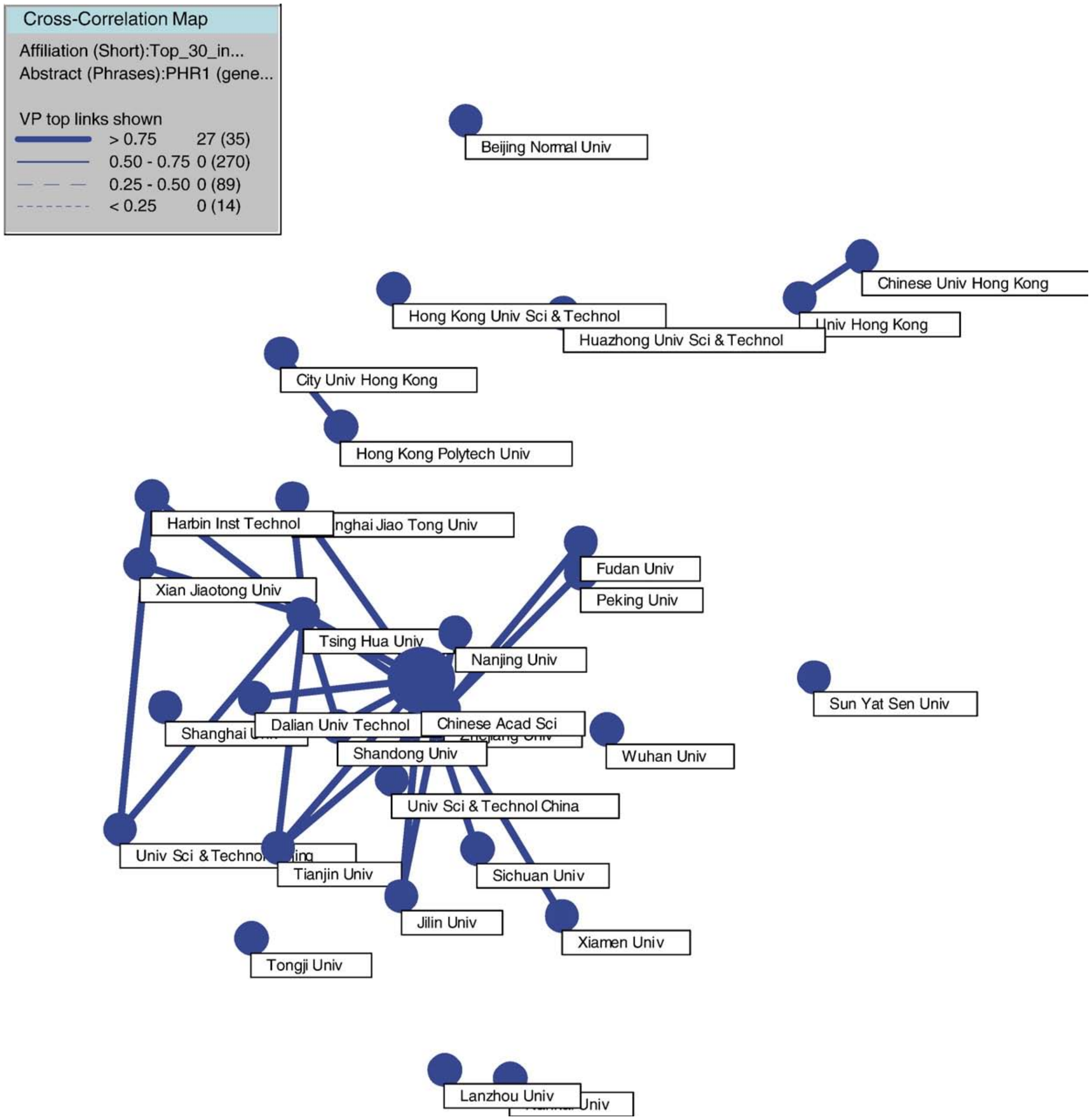

Fig. 5. Chinese Institution-phrase cross-correlation map (2005) including generic phrases.

Fig. 6 are from Hong Kong, suggesting the remnants of some degree of technical or thematic independence of Hong Kong from the more traditional Mainland institutions. Phrases associated with many of the institutions in this group are in the areas of chemistry and, in particular, items related to nanotechnology (e.g., X-ray diffraction, microstructure, SEM). A medical group associated with the University of Hong Kong and Chinese University of Hong Kong is clearly evident (e.g., infection, SARS, genes).

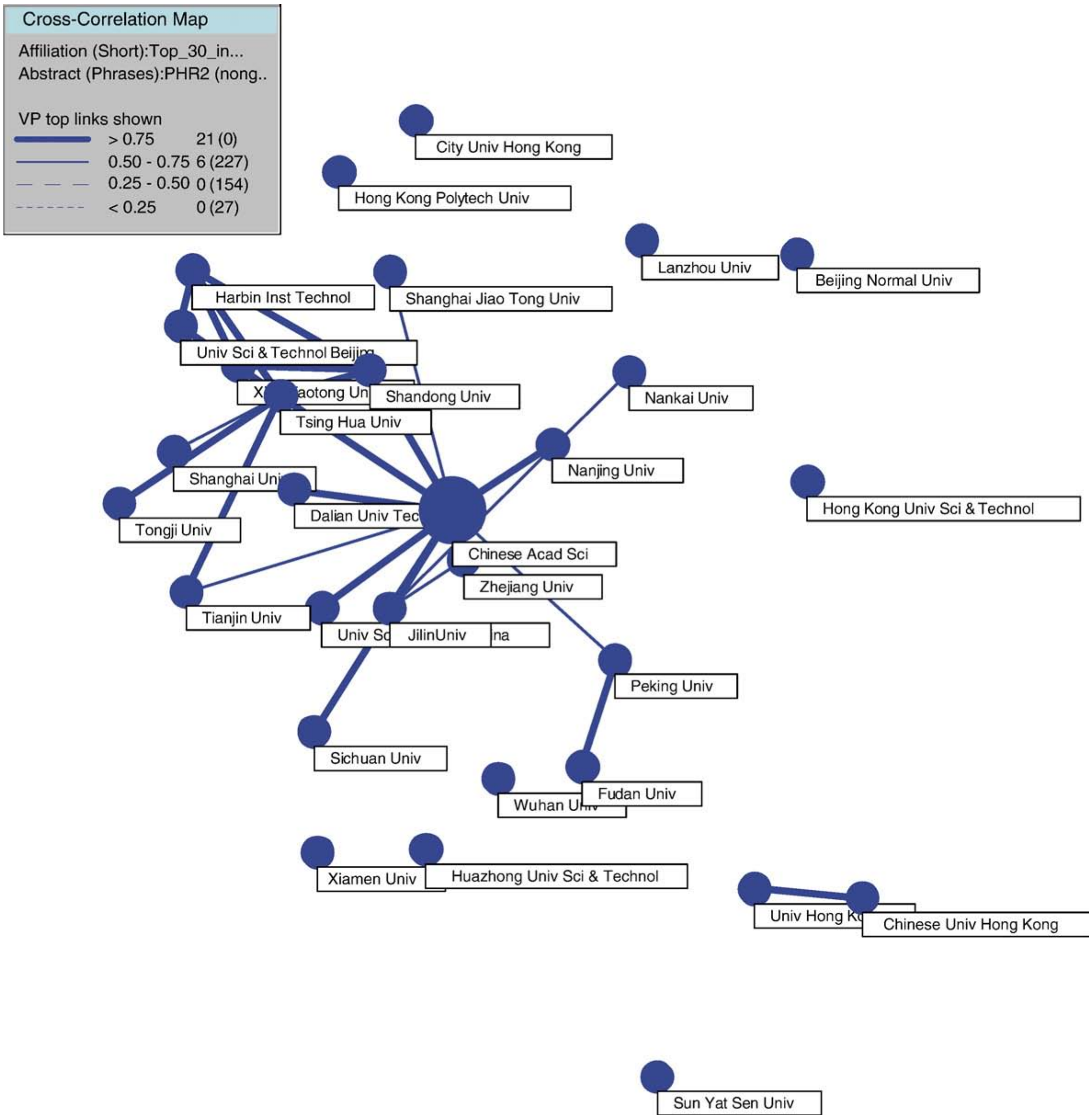

Fig. 6. Chinese Institution-phrase cross-correlation map (2005) with generic phrases removed.

2.2. Collaborative countries

In November 2005, the SCI was accessed to identify the main collaborating countries with China on research articles, in the period 2004–2005. The results are as follows. The format is the name of the country, followed by the number of articles that contained at least one country author and one Chinese author.

Table 8
Top ten collaborations between China and other countries for selected years

1980		1985		1990		1995		2000		2005	
Country/ territory	Record count	Country/ territory	Record count	Country/ territory	Record count	Country/ territory	Record count	Country/ territory	Record count	Country/ territory	Record count
People's R China	692	People's R China	3105	People's R China	6991	People's R China	11397	People's R China	29294	People's R China	72362
USA	37	USA	365	USA	644	USA	914	USA	2411	USA	5995
France	11	Japan	76	Japan	191	Japan	377	Japan	1082	Japan	2411
Denmark	8	England	56	Germany	156	Germany	309	Germany	694	Germany	1422
Fed Rep Ger	5	Fed Rep Ger	54	Canada	130	England	227	England	596	England	1401
Netherlands	4	France	48	England	122	Hong Kong	221	Canada	418	Canada	1175
Sweden	3	Canada	37	France	96	Canada	198	Australia	382	Australia	1024
England	2	Australia	25	Italy	84	France	174	France	325	France	866
Belgium	1	Sweden	20	Australia	47	Italy	141	Singapore	299	Singapore	799
Canada	1	Italy	12	Hong Kong	38	Australia	109	Taiwan	229	South Korea	712
Hong Kong	1	Netherlands	12	Sweden	38	Sweden	96	Italy	227	Taiwan	474

Table 9
Citations of papers with Chinese authors for selected years

Year	Number of papers	Papers with more than 100 cites	Number of cites		
			Median of top 20	Median of top 1%	Overall median
1980	692	3	37	96	1
1985	3105	20	172	131	1
1990	6991	36	209	101	2
1995	11,397	42	195	85	2
2000	29,294	63	199	70	3
2005	72,362	2	23	11	0

Note: the data for 2005 reflects the short period of time elapsed since the 2005 papers were available to be cited.

China (118659); USA (9919); Japan (4247); Germany (2450); England (2295); Canada (1923); Australia (1811); France (1374); Singapore (1334); South Korea (1197); Taiwan (870); Russia (651); Italy (632); Sweden (626); India (623).

To determine the citation impact of collaboration, two cases were compared. The first case consisted of all research articles in the SCI published from 1995–1999 having at least one author with a Peoples Republic of China address. The second case consisted of all research articles in the SCI published from 1995–1999 with China authors only.

The first case (China and collaborators) produced the following results:

- Articles retrieved, 83689;
- Median citations of total articles retrieved, 2;
- Median citations of top ten cited articles retrieved, 604;
- Median citations of top 5% articles retrieved, 35.

The second case (China only) produced the following results:

- Articles retrieved, 62018;
- Median citations of total articles retrieved, 2;

Table 10A
Journals containing most cited papers (>150 citations) published in 1979–1987

Frequency	Most cited — 1979–1987
	Journals
5	Nature
4	Physical Review Letters
2	Circulation
2	International Journal of Cancer
2	Journal of Mathematical Analysis and Applications
2	New England Journal of Medicine
2	Zeitschrift fur Physik C-Particles and Fields

Table 10B
Journals containing least cited papers (zero citations) published in 1979–1987

Frequency	Least cited — 1979–1987
	Journals
6	Kexue Tongbao
5	Chinese Physics
3	Acta Geophysica Sinica
3	Acta Pharmacologica Sinica
3	Acta Zoologica Sinica
3	Chinese Medical Journal
3	Scientia Sinica Series A-Mathematical Physical Astronomical & Technical Sciences
3	Scientia Sinica Series B-Chemical Biological Agricultural Medical & Earth Sciences

- Median citations of top ten cited articles retrieved, 239;
- Median citations of top 5% articles retrieved, 25.

Thus, approximately one-quarter of research articles having at least one author with a China address were the result of China's collaboration with other countries. The impact of collaboration was negligible on median citations of the total. The impact of collaboration was substantial on the top ten cited articles, and was noticeable on the top 5% of cited articles.

To determine the main technical areas of collaboration, two cases were examined: USA and Japan. The 2000 most recent articles for USA–China papers and for Japan–China papers (in late 2005) were downloaded from the SCI/SSCI. A phrase frequency analysis of the Abstracts was performed for each country combination, the highest frequency high technical content phrases were extracted, and the phrases were ordered by frequency of occurrence.

Table 10C
Journals containing most cited papers published in 1998–2003

Frequency	Most cited — 1998–2003
	Journals
15	Lancet
13	Physical Review Letters
11	Nature
9	New England Journal of Medicine
9	Science
8	Applied Physics Letters
4	Physics Letters A
3	Angewandte Chemie-International Edition
3	Journal of the American Chemical Society
3	Nature Cell Biology
3	Nature Genetics
3	Nucleic Acids Research
3	Physics Letters B

Table 10D
Journals containing least cited papers published in 1998–2003

Frequency	Least cited — 1998–2003
	Journals
5	Applied Mathematics and Mechanics—English Edition
5	Chinese Journal of Analytical Chemistry
4	Chinese Chemical Letters
4	Communications in Theoretical Physics
3	Chemical Journal of Chinese Universities—Chinese
3	Chinese Medical Journal
3	Chinese Science Bulletin
3	High Energy Physics and Nuclear Physics—Chinese Edition
3	Journal of Inorganic Materials
3	Progress in Natural Science
3	Science in China Series A—Mathematics Physics Astronomy
2	Acta Petrologica Sinica
2	Chinese Journal of Chemical Engineering
2	Ferroelectrics
2	International Journal of Pressure Vessels and Piping
2	Journal of Chemical Research—S
2	Journal of Materials Processing Technology
2	Journal of Materials Science & Technology
2	Journal of Rare Earths
2	Spectroscopy Letters
2	Statistics & Probability Letters
2	Transactions of Nonferrous Metals Society of China

The two areas that stand out for both collaborative groups (China–USA; China–Japan) are biomedical and nanotechnology. However, when frequencies of similar phrases from each group are taken into account, for the China–USA articles, biomedical comes first and nanotechnology second. For the China–Japan articles, nanotechnology ranks higher relative to biomedical. Given China's relative (to the USA) investment strategy emphasis in nanotechnology, as will be shown later, and lesser relative investment emphasis in biomedical, *the collaborative research relationship with Japan appears to be more quid pro quo than is the relationship with the USA.*

The top ten collaborative countries with China as a function of time are shown in Table 8. For the past 25 years, the USA has remained China's largest collaborator. Additionally, the level of collaboration by the USA has remained relatively constant at about one-tenth of the total number of Chinese publications. By contrast, Table 5 showed a much larger fraction of collaborators involved with higher Impact Factor journals. Therefore, the USA makes a much larger contribution to collaboration on the high Impact Factor journals. In Table 8, some of the Northern and Central European countries that appeared at the earlier times were replaced by East Asian countries for the latest dates.

2.3. Citation bibliometrics

2.3.1. Citation statistics on journals

The second group of metrics assessed was counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics [6], much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers [7,8].

2.3.2. Citation trends over time

The numbers of citations of papers with at least one Chinese author are presented in Table 9. There has been a relatively steady increase in the numbers of papers with greater than 100 citations. However, there has also been an exponential increase in the total number of papers, with the fraction of papers having 100 or more citations dropping steadily. While the median citations of the top twenty papers have essentially remained constant since 1990, the median citations of the top 1% have dropped about 30%. In the latter case, the number of papers in the top 1% increased from 70 in 1990 to 293 in 2000. At the high end, the quality distribution did not scale proportionally to growth, a problem that appears to plague almost every rapid growth situation. Overall, the median increased, reflecting an improvement at the low-medium sector.

2.4. Characteristics of most cited vs. least cited papers

The papers with the most citations (greater than 150) were compared with those with the least citations (zero) for two separate time periods (1979–1987, 1998–2003), in order to track the changes in characteristics of these papers. The papers with zero citations were obtained by random sampling of all articles with zero citations published in the time frame of the most cited articles. The results are presented in Tables 10A, 10B, 10C, 10D, 11A, 11B, 11C, 11D, 12A, 12B, 12C, 12D, 13A, 13B, 13C, 13D.

Table 11A
Authors producing most cited papers published in 1979–1987

Frequency	Most cited — 1979–1987
	Authors
3	Chen, GM
3	Tapponnier, P
2	Abramowicz, H
2	Armijo, R
2	Avolio, AP
2	Burg, JP
2	Degroot, JGH
2	Dydak, F
2	Eisele, F
2	Flottmann, T
2	Geweniger, C
2	Guyot, C
2	Han, TL
2	He, JT
2	Heller, F
2	Klasen, HP
2	Kleinknecht, K
2	Knobloch, J
2	Krolikowski, J
2	Lt, TD

Table 11B
Authors producing least cited papers published in 1979–1987

Frequency	Least cited — 1979–1987
	Authors
2	Zhou, ZY
1	Bao, ZT
1	Bei, MZ
1	Chaohao, G
1	Chen, B
1	Chen, C
1	Chen, P
1	Chen, YZ
1	Chen, Z
1	Cheng, GB
1	Cheng, JG
1	Cheng, S
1	Chu, XX
1	Dei, XJ
1	Ding, SY
1	Ding, XM
1	Dong, GS
1	Dong, ZH
1	Duan, HM
1	Fan, QI

Table 10A displays the highest frequency journals with Chinese-authored articles published between 1979 and 1987 and having greater than 150 citations, while Table 10B displays the journals containing the zero cited articles. Tables 10C and 10D contain the same type of data for 1998–2003 articles. Tables 11A, 11B, 11C, 11D, 12A, 12B, 12C, 12D, 13A, 13B, 13C, 13D show the authors, institutions, and collaborative countries, respectively, associated with Chinese-authored papers having greater than 150 citations and those having zero citations.

The journals containing the most cited papers from 1979–1987 (Table 10A) are well-known international journals, while those containing the least cited papers (Table 10B) are domestic Chinese journals. Interestingly, both groups of journals cover broadly similar topical material (physics, medicine, mathematics), with the least cited group also including geophysics, pharmacology, and zoology. For 1998–2003, the most cited journals (Table 10C) are the well-known international journals covering multi-disciplinary science, medicine, physics, chemistry, and biology. Again, the highest frequency least cited journals (Table 10D) are almost exclusively domestic Chinese journals, with still a strong focus on mathematics, chemistry, physics, and materials. Publishing in the domestic Chinese journals is not the path to high citations.

Of the authors of the most highly cited papers published between 1979 and 1987 (Table 11A), about 80% of the names are non-Asian. For the least cited (Table 11B), all the names listed are Chinese. For the most cited in the 1998–2003 time frame (Table 11C), half the names are now Asian. Spot checks showed these Chinese names have China addresses. Consider the first two names listed, for example, J. Zhang and K.Y. Yuen. Zhang is a participant in a large international high energy physics experiment, with over a hundred co-authors. Yuen is part of a Chinese research team that addresses national health issues such as

Table 11C
Authors producing most cited papers published in 1998–2003

Frequency	Most cited — 1998–2003
	Authors
9	Zhang, J
7	Yuen, KY
6	Abe, K
6	Emery, S
6	Hu, H
6	Kang, JH
6	Kim, HJ
6	Lebedev, A
6	Lim, W
6	Matsumoto, T
6	Peiris, JSM
6	Tanaka, Y
6	Watanabe, Y
5	Batignani, G
5	Bettarini, S
5	Bozzi, C
5	Calderini, G
5	Carpinelli, M
5	Charles, E
5	Cowan, G

SARS and influenza virus. For the least cited in this time frame (Table 11D), all the names are again Chinese.

Table 12A displays institutions generating the highly cited papers between 1979 and 1987. Most of the top institutions are Chinese, although for the results to be compatible with the large number of non-Asian names in this group, there must be many non-Chinese institutions appearing once. The few institutions shown with unit frequency support this statement. The least cited institutions in this time frame (Table 12B) are all Chinese, in full compatibility with the author name ethnicity results.

The most cited institutions from 1998–2003 (Table 12C) are roughly half Chinese/half USA. This is graphic evidence of the contribution of USA researchers to highly cited papers with Chinese authors. Further, the two Hong Kong universities listed have almost twice the number of highly cited papers as the Chinese Academy of Science, even though (from Table 6) the Chinese Academy of Science had four times the number of publications as these two Hong Kong institutions combined. In fact, the two most productive Mainland Chinese universities from Table 6, Tsing Hua University and Zhejiang University, do not show up on Table 12C, above the threshold listed. However, these two Mainland Chinese universities do show up on the least cited list Table 12D, along with the Chinese Academy of Science.

In the list of collaborating countries associated with highly cited papers (Table 13A), the USA predominates. The USA is represented in 45% of the highly cited papers, yet according to Table 8 data from that time frame, the USA is represented in about 10% of the total papers, give or take a few percent. Thus, the USA representation on the highly cited papers is out of all proportion to its total representation. As was evident from the author and institution results, the least cited papers in this time frame are all from China alone.

Table 11D
Authors producing least cited papers published in 1998–2003

Frequency	Least cited — 1998–2003
	Authors
3	Liu, Y
2	Chen, J
2	Chen, YZ
2	Li, GH
2	Li, L
2	Liu, YC
2	Liu, YS
2	Pan, W
2	Shen, JC
2	Wang, M
2	Wang, QY
2	Wang, SL
2	Wang, YX
2	Yang, J
2	Yang, XQ
2	Zhang, JY
2	Zhang, RF
2	Zhang, Y
2	Zheng, HP
2	Zhu, ZM

In 1998–2003, half the highly cited papers have USA representation, while only 2% of the least cited papers have USA representation. Even though many countries are on the least cited list, their total frequency is small, and China predominates.

2.4.1. Most cited journals

Approximately 2000 journals were cited ten or more times. The top twenty most cited journals are listed below in Table 14. The most cited journals appear to be primarily English language journals, in contrast to many of the journals containing most Chinese papers being Chinese journals. This suggests that in the 2005 time frame there may be a larger dependence on English Language (i.e. foreign) journals than on China's own internal journals, at least for Chinese papers published in journals accessed by the SCI.

The median Impact Factor of nineteen of the twenty journals listed in Table 14 (one journal did not have an Impact Factor listed) is 5.16. This is contrasted with the median Impact Factor of 23 of the 25 journals containing the most papers and listed in Table 4 (0.82). This almost order of magnitude difference in Impact Factor (between the journals in which the Chinese researchers publish and the journals that they reference) indicates Chinese researchers may not be publishing in the highest research impact journals. Since Impact Factor is discipline dependent, a discipline-based comparison of Tables 4 and 14 may be instructive.

The median of the Impact Factors of the seven physics journals in Table 4 is 1.27, whereas the median of the Impact Factors of the seven physics journals in Table 14 is 4.31, a factor of ~3.5 difference. The median of the Impact Factors of the seven chemistry journals in Table 4 is 0.77, whereas the median of the

Table 12A
Institutions producing most cited papers (>150 citations) published in 1979–1987

Frequency	Most cited — 1979–1987
	Institutions
11	Chinese Acad Sci
8	Acad Sinica
5	Chinese Acad Med Sci
5	Inst High Energy Phys
3	CENS
3	Minist Geol
2	Beijing Med Coll
2	Chinese Acad Geol Sci
2	Plast Surg Hosp
2	Univ Dortmund
2	Univ Heidelberg
2	Univ Paris 07
1	Baker Med Res Inst
1	Beijing Normal Univ
1	Beijing Univ
1	Cardiovasc Inst
1	Childrens Hosp Med Ctr
1	China Cent Inst Nationalit
1	China Natl Ctr Prevent Med
1	Chinese Acad Agr Sci
1	Chinese Univ Sci & Technol
1	Christie Hosp & Holt Radium Inst
1	Concordia Univ
1	DESY
1	Emory Univ

Table 12B
Institutions producing least cited papers published in 1979–1987

Frequency	Least cited — 1979–1987
	Institutions
9	Chinese Acad Sci
8	Acad Sinica
5	Univ Sci & Technol China
4	Fudan Univ
3	Beijing Univ
2	Suzhou Med Coll
1	Acad Mil Med Sci
1	Acad Sci Agr China
1	Beijing Agr Univ
1	Beijing Polytech Univ
1	Chinas Environm Sci Acad
1	Chinese Acad Med Sci
1	E China Inst Chem Technol

Table 12C
Institutions producing most cited papers published in 1998–2003

Frequency	Most cited — 1998–2003
	Institutions
19	Chinese Acad Sci
19	Univ Hong Kong
15	Chinese Univ Hong Kong
14	Inst High Energy Phys
12	Acad Sinica
11	Peking Univ
11	Univ. Calif Berkeley
10	Univ. Sci & Technol China
10	Univ. Texas
10	Univ. Washington
9	Yale Univ
8	Univ. Wisconsin
8	Vanderbilt Univ

Impact Factors of the seven chemistry journals in Table 14 is 3.46, a factor of ~7 difference. The median of the Impact Factors of the eight materials journals in Table 4 (including two with no Impact Factor listed) is 0.35, whereas the Impact Factor of the one materials journal in Table 14 is 1.71, a factor of ~5 difference. The one general science journal in Table 4 has an Impact Factor of 0.78, whereas the three general science journals in Table 14 have a median Impact Factor of 31.86, a factor of ~40 difference. The one medical journal in Table 4 has an Impact Factor of 0.56, while the one biology journal in Table 14 has an Impact Factor of 6.36.

While these comparisons are for the top 20–25 journals only, and the Impact Factors have not been weighted by the numbers of papers in each journal, it is quite clear that, on average, the Chinese researchers are not publishing extensively in the high research impact journals they are referencing.

Table 12D
Institutions producing least cited papers published in 1998–2003

Frequency	Least cited — 1998–2003
	Institutions
26	Chinese Acad Sci
9	Tsing Hua Univ
7	Acad Sinica
7	Shandong Univ
7	Univ Sci & Technol China
6	Univ Hong Kong
5	Peking Univ
4	Dalian Univ Technol
4	Fudan Univ
4	Jilin Univ
4	Lanzhou Univ
4	Nanjing Univ
4	Zhejiang Univ

Table 13A
Countries producing most cited papers (>150 citations) published in 1979–1987

Frequency	Most cited — 1979–1987
	Countries
57	Peoples R China
26	USA
10	France
8	Fed Rep Ger
6	Switzerland
4	England
3	Australia
2	Canada
2	Japan
1	Austria
1	India
1	Italy
1	Netherlands
1	Pakistan

However, as was shown previously, they are increasing their (presently small) contribution to high Impact Factor journals.

2.5. Research investment strategy analysis

Understanding China's research investment allocations is the first step to understanding their research investment strategy. This understanding provides context for specific technical thrusts identified. Two approaches were used to determine China's research investment allocations. Both are computational linguistics based, with the first based on document groupings, and the second based on phrases.

The first approach uses document clustering to generate technical categories and the numbers of documents that populate these categories. This approach does not use funding allocations directly, but rather the proxy metric of published documents as a reflection of levels of effort.

In partitional clustering, the number of clusters desired is input, and all documents in the database are included in those clusters. Clustering was conducted for the 2004–2005 documents retrieved from the SCI/SSCI. There were 256 clusters run for the retrieved articles, and these clusters are listed in [9], in the order by which they appear on the hierarchical tree. They were aggregated into a hierarchical taxonomy using a hierarchical tree generated by the CLUTO software.

Table 13B
Countries producing least cited papers (zero citations) published in 1979–1987

Frequency	Least Cited— 1979–1987
	Countries
57	Peoples R China

Table 13C
Countries producing most cited papers published in 1998–2003

Frequency	Most cited — 1998–2003
	Countries
136	Peoples R China
69	USA
29	England
28	Germany
23	France
16	Japan
15	Russia
14	Italy
11	Canada
11	Netherlands
11	South Korea
11	Taiwan

Fig. 7 contains the first four levels of the hierarchical taxonomy for 2004–2005 SCI records for China. Each cell in the matrix represents a technical category. There are four columns in Fig. 7, each column representing a level of the hierarchical taxonomy. The highest level (1) is the left-most column, and the lowest level (4) is the right-most column. The number in parenthesis in the cell is the number of records assigned by the algorithm to the category.

In the following discussion, the categories of Levels 1 and 4 in Fig. 7 will be described. The contents of the categories in Levels 2 and 3 are self-evident from their headings and from the contents of Levels 1 and 4.

Table 13D
Countries producing least cited papers published in 1998–2003

Frequency	Least cited — 1998–2003
	Countries
136	Peoples R China
4	Germany
3	Hong Kong
2	Australia
2	Canada
2	England
2	France
2	Japan
2	USA
1	Mexico
1	North Ireland
1	Russia
1	Singapore
1	Spain
1	Taiwan

Table 14
Most cited journals

Journal	#Papers	Impact Factor	Theme
Phys Rev Lett	2592	7.22	Physics
J Am Chem Soc	2196	6.9	Chemistry
Nature	2191	32.18	Science
Phys Rev B	2027	3.08	Physics
Science	1995	31.86	Science
Appl Phys Lett	1737	4.31	Physics
J Appl Phys	1433	2.26	Physics
J Chem Phys	1174	3.11	Chemistry
P Natl Acad Sci USA	976	10.45	Science
Anal Chem	924	5.45	Chemistry
J Biol Chem	917	6.36	Biology
Phys Rev D	834	5.16	Physics
Phys Rev A	779	2.9	Physics
Inorg Chem	757	3.45	Chemistry
J Phys Chem—US	738		Physics
J Am Ceram Soc	738	1.71	Materials
Macromolecules	714	3.9	Chemistry
Angew Chem Int Edit	687	9.16	Chemistry
Astrophys J	641	6.24	Physics
J Org Chem	612	3.46	Chemistry

Level 1 is divided into two categories: (a) Physical and Materials Sciences (16030) and (b) Medical, Mathematics and Modeling, Life Sciences (18969). Physical and Materials Sciences covers research in physics, chemistry, and materials, and it places a strong emphasis on applied chemistry; catalysis; the structure of materials and molecules; preparation of materials (especially alloys and ceramics); the performance of metals, alloys, composites, and ceramics; and the preparation and application of thin films.

Medical, Mathematics and Modeling, Life Sciences is focused on the modeling of systems and phenomena; understanding the underlying mathematical theory related to the models; computational techniques and networking; agricultural research; and medical research and treatment of diseases.

Level 4 is divided into thirteen categories. They are described in order of their listing in Fig. 7, starting from the top. Metrics for each cluster are included in the computer output, and where significant, are discussed.

- Analytical Chemistry Methods, Spectral Analyses (3277).
  This category has two major thrusts: (a) applications of carbon nanotubes and nanotechnology (596) and (b) applied chemical separations, chemical identification, and chemical sensors (2681). The USA is the major collaborator, with Japan as a strong contributor.
- Catalytic Reactions (1935).
  This category is focused on catalytically-promoted chemical reactions and synthesis. It appears to be primarily a fundamental research category looking at topics such as chemical kinetics and thermodynamics, and understanding reaction mechanisms. The USA and Japan are the primary collaborators, both contributing at approximately an equal level.

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
physical and materials sciences (16030)	chemistry, molecular structure (7693)	chemical reactions, catalysis (5212)	analytical chemistry methods, spectral analyses (3277)
			catalytic reactions (1935)
		crystal and chemical structure, synthesis (2481)	
	materials and structural properties, thin films, ceramics (8337)	performance of metals, alloys, composites, ceramics (6564)	powders, microcrystalline materials, crystal growth, materials preparation (3748)
			metal alloys, ceramics (2816)
		thin film preparation, properties, and applications (1773)	
medical, mathematics and modeling, life sciences (18969)	numerical modeling and simulation (11607)	mathematical and numerical methods, theoretical equations and models (6957)	modeling of systems and phenomena (4871)
			mathematical theory and application (2086)
		computational processing techniques, networks, information processing (4650)	
	biochemistry and biophysics, cellular and genetic biology (7362)	health research, disease treatment, plant and environmental science (3649)	soils, plants, geoscience (2023)
			clinical treatment of diseases, occupational health, cancer research and treatment (1626)
		biochemistry and molecular biology, gene expression, pharmacology (3713)	gene expression and sequencing (1473)
			cell expression, pharmacology (7362)

Fig. 7. Taxonomy of SCI 2004–2005 Chinese research articles.

- Crystal and Chemical Structure, Synthesis (2481).
  This category is focused on the synthesis of crystals and chemical compounds and on understanding their fundamental properties. Much of this work is fundamental research, however there is some indication of applications such as biodetection and detection of various chemicals. Unexpectedly, Malaysia is the strongest collaborator in this area, with the USA and Japan following closely behind.

- Powders, Nanocrystalline Materials, Crystal Growth, Materials Preparation (3748)
  This category is focused on understanding microstructural properties of materials and the growth of crystals. The work appears to be primarily basic research. The USA and Japan are the strongest collaborators.
- Metal Alloys, Ceramics (2816).
  This category has two major thrusts: (a) Metal alloy performance and characterization (especially with respect to deformation, corrosion, temperature, composition, and preparation technique) (1301) and (b) ceramic properties, performance, and preparation (1515). Japan is the main collaborator, followed closely by the USA.
- Thin Film Preparation, Properties, and Applications (1773).
  This category focuses on the preparation of thin films, understanding properties, and applications (e.g., sensors, detectors, lubricants, and electronics). The USA and Japan are the largest collaborators.
- Modeling of Systems and Phenomena (4871).
  This category has two major thrusts: (a) laser optics and applications (1786) and (b) modeling of engineering performance (e.g., hydrodynamics, stress, flow, heat transfer) (3085). The USA is the largest collaborator, more than doubling the efforts by Japan and Germany.
- Mathematical Theory and Application (2086).
  This category focuses primarily on fundamental mathematics with some applications to specific problems. The USA is the largest collaborator.
- Computational Processing Techniques, Networks, Information Processing (4650).
  This category is focused on the development of algorithms and models, and their computational implementation to solve applied problems (e.g., robotics, logistics/optimization, automation, command and control, visualization). The USA is by far the largest collaborator.
- Soils, Plants, and Geoscience (2023).
  This category focuses on agriculture research, plant biology, and understanding geoscience (e.g., earthquakes). The USA is the largest collaborator, more than doubling the collaborative output by Japan.
- Clinical Treatment of Diseases, Occupational Health, Cancer Research, and Treatment (1626).
  This category focuses on the clinical treatment of diseases and other medical research, with a significant effort devoted to cancer research. The USA is the largest contributor.
- Gene Expression and Sequencing (1473).
  Major topics in this category include virus detection, DNA sequencing, and gene expression. The USA is a major collaborator.
- Cell Expression, Pharmacology (2240).
  This category focuses on cell expression, laboratory animal tests for medical purposes, and the development of drug treatments. The USA is the largest collaborator.

In summary, the USA is a major contributor on almost every cluster, with Japan being a strong second. Malaysia was the strongest collaborator in the ‘crystal and chemical structure, synthesis’ cluster. Further analysis showed that about 85% of Malaysia’s co-authored papers with China occurred in this category.

2.6. Structure of Chinese technology in technical categories

A process similar to partitional clustering was used to generate a hierarchical taxonomy on the 2003 EC data. The first level of the technology taxonomy has two categories: Computer Sciences (4721 records)

Table 15
Chinese strengths — SCI/SSCI

Country	DMWord	# of China Abstracts	# of USA Abstracts	Absolute ratio	Normalized ratio
China	Linear matrix	144	25	5.76	21.8448
China	XRD	1989	404	4.923267	18.67149
China	Light irradiation	101	22	4.590909	17.41102
China	Spark plasma sintering	55	12	4.583333	17.38229
China	UV–vis* absorption	174	41	4.243902	16.095
China	Periodic solution*	222	63	3.52381	13.36405
China	HRTEM	212	62	3.419355	12.9679
China	Photocatal*	353	114	3.096491	11.74344
China	Impact strength	65	22	2.954545	11.20511
China	Nanobelt*	147	53	2.773585	10.51882
China	Corrosion resist*	161	62	2.596774	9.848266
China	Multi-walled carbon nanotube*	84	34	2.470588	9.369706
China	Thermogravimet*	432	190	2.273684	8.622947
China	Nanorod*	438	201	2.179104	8.264254
China	Wurtzite	151	77	1.961039	7.43724
China	Photoluminescen*	911	493	1.84787	7.008048
China	Photo degradation	88	49	1.795918	6.81102
China	Alloying	154	87	1.770115	6.713161
China	Fluorescen* Spectr*	407	236	1.724576	6.540456
China	Scanning electron microscop*	1466	864	1.696759	6.434959
China	PZT	127	75	1.693333	6.421967
China	CEO2	105	64	1.640625	6.22207
China	Flexural strength	72	47	1.531915	5.809787
China	Visible light	175	121	1.446281	5.485021
China	Fuzzy cluster*	35	27	1.296296	4.916204
China	Luminescen*	587	485	1.210309	4.590098
China	Optical microscop*	239	203	1.17734	4.465062
China	Nonlinear optic*	209	181	1.154696	4.379185
China	Spectrophotomet*	298	270	1.103704	4.185796
China	Wavelet transform*	128	117	1.094017	4.14906
China	Nanostructur*	803	910	1.13325	3.346293

and Physical Sciences (5228 records). Percentage-wise, this is a split of 47/53%. The second taxonomy level is generated by sub-dividing each first level category by two. Computer Sciences divides into Cybernetics & Systems Engineering (3902) and Signal Processing (819), while Physical Sciences divides into Materials Science (3477) and Chemistry & Nanotechnology (1751). The lower taxonomy levels are generated in the same manner as above. In the fourth taxonomy level, several categories stand out as receiving significantly more focus than the others. These categories are Systems Theory (23.4%) and Structural Mechanics & Materials (20.1%) with the most focus, followed by Applied Measurements (9.3%), Power/Energy Market Enterprises (8.6%), and Organic Chemistry (7.2%) as compared to the other eleven categories ranging from 1.3–4.9%.

The Abstracts also cover a broad range of fields ranging from industrial to high tech electronics that are indicative of a large society growing to sustain itself and become technologically competitive on a global scale. Examples of some key areas receiving emphasis are as follows; Energy/Power Generation, Mining,

Table 16
USA strengths — SCI/SSCI

Country	DMWord	# of China Abstracts	# of USA Abstracts	Absolute ratio	Normalized ratio
USA	Physician*	33	2790	84.54545	22.29464
USA	"Drug abuse" or "substance abuse"	13	816	62.76923	16.55225
USA	Antiretroviral	15	784	52.26667	13.78272
USA	Counseling	13	673	51.76923	13.65155
USA	Disabilit* or disabled	35	1698	48.51429	12.79322
USA	Medication*	54	2553	47.27778	12.46715
USA	Anxiety	34	1481	43.55882	11.48646
USA	Poverty	11	474	43.09091	11.36307
USA	Psoriasis	5	165	33	8.7021
USA	Intervention*	247	7625	30.87045	8.140536
USA	Traumatic	33	998	30.24242	7.974927
USA	Mental	100	2922	29.22	7.705314
USA	Pathophysiology	35	938	26.8	7.06716
USA	Fear	23	608	26.43478	6.970852
USA	Life expectancy	7	178	25.42857	6.705514
USA	Sleep	37	878	23.72973	6.25753
USA	Racial or racism or ethni*	109	2561	23.49541	6.19574
USA	Clinical trial	66	1546	23.42424	6.176973
USA	Whites or Caucasian	47	1027	21.85106	5.762126
USA	Infant* or neonatal	153	3306	21.60784	5.697988
USA	Feeling* or emotion*	107	2217	20.71963	5.463765
USA	Nicotine	24	495	20.625	5.438813
USA	Space telescop*	28	437	15.60714	4.115604
USA	Pain*	241	3636	15.08714	3.978478
USA	Epidemiology	126	1881	14.92857	3.936664
USA	Vascular disease	22	323	14.68182	3.871595
USA	Star formation*	31	417	13.45161	3.54719
USA	Pharmacolog*	181	2303	12.72376	3.355255
USA	Stellar	67	848	12.65672	3.337576
USA	Telescope	72	880	12.22222	3.223
USA	Galax*	161	1320	8.198758	2.162012

Materials & Structural Mechanics, Signal Processing, Systems Engineering, Transportation & Traffic flow, Robotics, Sensors & Diagnostics, Advanced Communications, Nanotechnology, Assessment Methods, Mathematics, Environmental & Ecological, Modeling & Simulation, and Control Theory. All of these areas have applications that can be of military significance.

Efforts in energy and power generation include hydroelectric, nuclear, and fossil fuels (such as coal), with the emphasis on the latter. Improvements are being sought for more efficient yields of energy from these resources. Power generation spans from Power Plants to vehicles to small electronic devices. The efforts in fossil fuels are closely tied with mining and structural developments.

The efforts in mining include identifying areas of opportunity for different resources, and improving mine structures to prevent collapse. These efforts can be closely associated with other work in remote sensing to help locate resources and conduct environmental impact studies. The same efforts to improve structural developments in mines might also be applied to underground facilities. Materials and structural mechanics

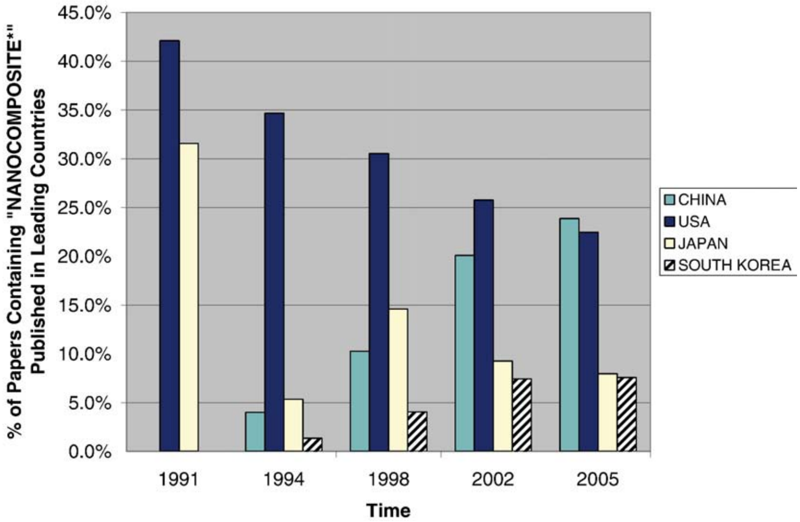

Fig. 8. Nanocomposite publication fraction vs. time.

fields range from the macro level (geologic formations and superstructures) to the micro and nano level (e.g. particles, ligands, compounds, films, and nanowires). There are specific references of structural analyses being done for a *New-Concept Submarine* and *low noise torpedo*, as well as for solid rocket motors.

Systems, control theory, modeling, and simulation are closely associated with all other areas. They range from the macroscopic, such as improving trafficability movements of large vehicles, resources, people, and robotics to the microscopic, such as gene manipulation. They are being conducted for topics small and large in numbers, such as tracking and/or controlling Unmanned Aerial Vehicles (UAVs) in a dense air traffic environment. Vibrational analysis is being performed with specific applications to missile launches on naval ships. Signal processing techniques are also closely related to these fields as well and incorporate wavelets, digital signal processing and neural networks. Applications of these studies include remote detection and biometrics.

Assessments, testing, and diagnostic methods include studies of text mining, Transmission Electron Microscopy (TEM), X-ray Diffraction (XRD), Magnetic Resonance Imaging (MRIs), and other high precision diagnostic instrumentation that can be used in high-yield weapons development. Long range plans are made that include research, such as the specific reference to a new 5-year coal mining plan.

Communications related research focuses on topics such as fiber optics, optical communications in seawater, digital, wireless networks, mobile networks, millimeter waveguides, blind signature schemes in cryptography, and security protocols.

2.7. Comparison of China's and USA's research investment allocations

The second approach for determining China's research investment allocations uses specific queries derived from Abstract phrases. To place these results in context, the records associated with very specific Chinese efforts are compared with those of the USA. The specific approach is described in the Introductory paper to this Special Issue. The results are shown on Tables 15 and 16.

Table 17
Chinese strengths — EC

Query phrase	# 2005 EC abstracts		Absolute ratio EC
	China	USA	China/USA
Bearing capacity	145	12	12.08
XRD	2213	237	9.34
Microhardness	174	22	7.91
Photoelectric	86	13	6.62
Diesel Engine	152	23	6.61
Wavelet transform	338	54	6.26
Fiber Bragg grating	115	19	6.05
Wear resistance	213	37	5.76
Annealing temperature	214	39	5.49
Impact strength	92	19	4.84
Magnetron	285	60	4.75
Countermeasures	57	13	4.38
Intrusion detection	100	23	4.35
Missile	100	24	4.17

The difference in thematic emphasis between the USA and China is dramatic! China emphasizes the physical and engineering sciences that underpin defense and commercial needs. The USA emphasizes research areas focused on medical, psychological, and social problems. There are even research areas where China leads the USA in absolute numbers of research articles published. In those areas, China's investment allocation relative to that of the USA is greater than four.

From Table 15, broad areas of China's emphasis include nanotechnology, luminescent phenomena, materials properties, and applied mathematics. A nanotechnology text mining study [2] showed that China

Table 18
USA strengths — EC

Query phrase	# 2005 EC abstracts		Absolute ratio EC
	China	USA	USA/China
Biochemistry	47	1498	31.87
Epithelial	9	182	20.22
C-terminal	17	308	18.12
Microbiology	13	196	15.08
Aeronautics	13	176	13.54
Transmembrane	14	176	12.57
Viral	10	121	12.1
Prostate	11	136	12.36
Cytoplasmic	13	162	12.46
Patient	28	351	12.54
Peptides	36	408	11.33
Transfection	9	101	11.22
Ecosystems	15	164	10.93
Mortality	13	127	9.77

was second to the USA in nanotechnology research article production. As the updated nanotechnology study [10] is showing (and which the present study is showing for other technical areas as well), at the next level or two lower in aggregation, there are nanotechnology sub-areas in which China is actually leading in absolute numbers of research article production. Moreover, as the updated nanotechnology study is showing, for those specific areas where China may be ahead of the USA or at parity with the USA, the growth gradients are strongly in China's favor and will widen the gap in the near future (see Fig. 8 for an illustrative example of this effect for nanocomposites).

The next two tables are similar to Tables 15 and 16, except that they contain common (to USA and China) high frequency phrases that were derived from the EC, instead of the SCI/SSCI. The EC data was generated in mid-October 2005.

Table 17 contains a set of phrases taken from the EC in which China had a large lead relative to the USA in terms of the ratio of record occurrences. In general, the EC is a much more applied database than the SCI/SSCI, and many of the words/phrases chosen in Tables 17 and 18 reflect that. Some of the phrases, such as XRD, were high frequency shared phrases not only in the China EC phrase list, but also in the China SCI/SSCI phrase list.

Analyses of Tables 17 and 18 confirm that in the EC, as in the SCI/SSCI, China's focus is on the physical and engineering sciences and especially engineering sciences, whereas the USA's relative focus is on health and biology-based research. In the overtly military-related terms (countermeasures, intrusion detection, missile), China has a significant presence. One interesting exception is the presence of 'aeronautics' in the list of USA dominant terms, similar to the presence of 'telescope' and 'galaxy' in the SCI/SSCI comparison. In technologies that require a large infrastructure, and therefore large investment, China has tended to be under-represented. That probably accounts for the 'aerospace' under-emphasis in the EC, and the 'telescope-galaxy' under-emphasis in the SCI/SSCI.

3. Conclusions

China's publication of SCI research articles has 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 multi-disciplinary science, medicine, and life science in 1980 to materials, chemistry, and physics in 2005, in that order. The major journals in which Chinese researchers publish have Impact Factors (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 access dates. Beyond increased productivity and recent sponsorship, recent SCI access by numerous journals must be considered as a contributor to the apparent high SCI growth rate of Chinese research articles. However, while the present median Impact Factor 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 numbers of Chinese articles published in very high Impact Factor journals, but produces negligible change in median citations of overall Chinese publications. Collaboration's major impact is at the high end.

The major Chinese research article producer, 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 collaborations with China–USA collaborations, the collaborative research relationship with Japan appears to be more quid pro quo than is the relationship with the USA, given the emphasis of the Japanese articles on nanotechnology first and biomedical second and the emphasis of the USA articles on biomedical first and nanotechnology second.

In comparing highly cited to poorly cited Chinese-authored documents, it was found that:

- Highly cited documents are published in the 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.
- 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 programs. In recent years, the Chinese participation in the most cited papers has increased to about half. The names associated with the least cited papers are all Chinese.
- In recent years, the institutions associated with the most cited papers are roughly half Chinese/half USA, showing the disproportionate contribution of USA 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).
- The USA is the predominant collaborator on the highly cited papers, having about five times the relative representation expected based on its collaboration with all China 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 thematic emphasis between the USA and China is dramatic! China emphasizes the physical and engineering sciences that underpin defense and commercial needs. The USA emphasizes research areas focused on medical, psychological, and social problems. There are even research areas where China leads the USA in absolute numbers of research articles published. In those areas, China’s investment allocation relative to that of the USA is greater than four. Moreover, when the time trends in these areas are examined, the much higher Chinese gradients relative to those of the USA imply that the gap will only widen in the future.

References

[1] 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: © Copyright THOMSON SCIENTIFIC® 2006. All rights reserved.

[2] R.N. Kostoff, J.A. Stump, D. Johnson, J.S. Murday, C.G.Y. Lau, W.M. Tolles, The structure and infrastructure of the global nanotechnology literature, J. Nanopart. Res. 8 (3–4) (2006) 301–321.

[3] M. Pecht, China’s Electronics Industry, William Andrew Publishing, NY, 2006.

[4] EC. Compendex®, Engineering Village, Elsevier Engineering Information, Inc., Hoboken, NJ 07030, 2006.

[5] Search Technology, Inc., Norcross, GA, (2006).

[6] E. Garfield, History of citation indexes for chemistry — a brief review, JCICS 25 (3) (1985) 170–174.

[7] R.N. Kostoff, The use and misuse of citation analysis in research evaluation, Scientometrics 43 (1) (1998) 27–43.
[8] M. MacRoberts, B. MacRoberts, Problems of citation analysis, Scientometrics 36 (3) (1996) 435–444.
[9] R.N. Kostoff, M.B. Briggs, R.L. Rushenberg, C.A. Bowles, M. Pecht, The Structure and Infrastructure of Chinese Science and Technology. DTIC Technical Report Number ADA443315, Defense Technical Information Center, Fort Belvoir, VA, 2006, http://www.dtic.mil/.
[10] R.N. Kostoff, R.G. Koytcheff, C.G.Y. Lau, Global nanotechnology research metrics, Scientometrics 70 (3) (2007) 565–601.

Ronald N. Kostoff received a Ph.D. in Aerospace and Mechanical Sciences from Princeton University in 1967. He has worked for Bell Laboratories, Department of Energy, and Office of Naval Research (ONR). He has authored over 100 technical papers, served as Guest Editor of three journal Special Issues, obtained two text mining system patents, and presently manages a text mining pilot program at ONR.

Michael B. Briggs, a Lieutenant Colonel in the United States Marine Corps Reserve, has a BS and MS in Nuclear Engineering, and an ME in Radiological Health Engineering from the University of Michigan. He is a registered Engineer-In-Training, and a member of the American Nuclear Society. He is presently serving in an active duty assignment with the Marine Corps Warfighting Laboratory supporting the Integration of Science and Technology into Marine Corps warfighting capabilities.

Robert L. Rushenberg has a BS in Mechanical Engineering and a minor in Astronomy from Iowa State University. He has worked in the field of Textual Data Mining for the past two and a half years, focusing on technology assessments and country science and technology. Most recently, Robert has worked on development and application of radical scientific discovery to improved water purification.

Christine A. Bowles has an MS in Materials Science and Engineering from Johns Hopkins University. Ms. Bowles has more than 15 years of experience in materials engineering with emphasis on corrosion and electrochemistry. She worked as a Materials Engineer at the Naval Surface Warfare Center (White Oak/Carderock) for almost 10 years, spent close to 4 years in the chemical processing industry, and is currently a Materials Engineer at DDL Omni Engineering.

Alan S. Icenhour received his M.S. and Ph.D. degrees in Nuclear Engineering from The University of Tennessee. He has 20 years experience with the nuclear fuel cycle, ranging from reactor operations to radiochemical research. He is presently a Technical Advisor on non-Proliferation at the National Nuclear Security Administration (NNSA), on rotational assignment from the Oak Ridge National Laboratory (ORNL), where he is a senior R&D staff member.

Kimberley F. Nikodym received a B.S. in Oceanography from the United States Naval Academy, and a dual MS in Oceanography and Meteorology from the Naval Post Graduate School. After serving in a number of command positions in the Navy, she left active duty in 1997. She has worked as a forecaster for the National Weather Service, and presently serves in the naval reserves supporting the U.S. Pacific Command.

Ryan B. Barth received his B. S. in Mechanical Engineering from Iowa State University. He currently works for DDL OMNI Engineering as a Mechanical Engineer. Mr. Barth has been performing Textual Data Mining on the transport and distribution logistics literature.

Michael Pecht has an M.S. and Ph.D. in Engineering Mechanics from the University of Wisconsin at Madison. He is a Professional Engineer, an IEEE Fellow, and an ASME Fellow. In addition to writing eighteen books on electronic products development, use and supply chain management, he has received the 3M Research Award for electronics packaging, the IEEE Undergraduate Teaching Award, and the IMAPS William D. Ashman Memorial Achievement Award for his contributions in electronics reliability analysis.