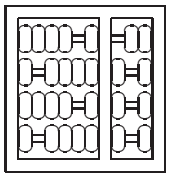

INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Scientific Production in Computer Science: Brazil and Other Countries

E. C. Xavier *Fábio Bezerra*
Jacques Wainer

Technical Report - IC-07-18 - Relatório Técnico

May - 2007 - Maio

Scientific Production in Computer Science: Brazil and Other Countries

E. C. Xavier Fábio Bezerra Jacques Wainer

May 31, 2007

Abstract

In this paper we present a study about scientific production in Computer Science in Brazil and several other countries, as measured by the number of articles in journals indexed by ISI. We compare the Brazilian production from 2001 to 2005 with some Latin American, Latin European, BRIC (Brazil, Russia, India, China), and other relevant countries. We also classify and compare these countries production according to the impact factor of the journals in which they were published, and according to each country known research and development investment.

The results show that Brazil has by far the largest production among Latin American countries, has a production about one third of Spain's, one fourth of Italy's and a little larger than Portugal's. The growth in Brazilian publications during the period places the country in the group with mid range growth, but regarding dollar productivity, Brazil joins the other BRIC countries as the ones with the lowest productivity. The distribution of Brazilian production according to impact factor is similar to most countries.

Key Words: Scientific Productivity, Science Metrics, ISI.

1 Introduction

This paper presents some results regarding the scientific production in Computer Science, comparing Brazil with several other countries, using the data available at the *Web of Science* [ISI 2006]. In most scientific areas, the journals indexed by the *Institute for Scientific Information* (ISI) [ISI 2006] are considered the most prestigious ones. That is only partially true in Computer Sciences - there are very well known and respected journals that are not yet indexed by ISI, for example ACM Journal on Experimental Algorithms, ACM Transactions on Algorithms, Journal of Discrete Algorithms, just to list a few in the area of algorithms.

Computer science has another peculiarity that not all scientific production is published in journals, but also in conferences and workshops. The computer science community has a strong respect for work published in some conferences, and has a long tradition of creating workshops to discuss cutting edge ideas and technologies.

There are other services and institutes that index Computer Science production. One example is *Citeseer* [Citeseer 2006] which not only indexes journals but also articles published in conferences, workshops and technical reports. Citeseer also calculates the impact

factor of the articles (and not of the journal as does ISI). But there are a few problems in using Citeseer to evaluate production. The first is that it is not yet clear how Citeseer selects the sources for the papers it indexes. Second, using the ISI indexed journals is the current practice scientometrics research, and thus our research using ISI can be compared with other research in other science areas; that would not be true if we used data from Citeseer, for example. Finally, Citeseer seems to have stop updating its impact factor list database early 2003.

Despite these characteristics of Computer Science (henceforth abbreviated as CS), we believe that a study of the production of Brazilian CS researchers in the ISI indexed journals is a first approximation to the whole of Brazilian CS production. To our knowledge this work is the first to measure Brazilian CS work extensively and to compare it with other countries.

Brazilian scientific production and productivity in other scientific areas has been studied in some previous work. Albuquerque et al. [Albuquerque et al. 2002] studied the regional aspects of scientific and technology productivity in Brazil. Leite [Leite 2006] used journals indexed by the ISI, to compare Brazil, China and India. In this case, he compared several areas (including, Mathematics, Computer Science, Physics, Biology, etc.) according to the percentage of published articles by each country and compared with each country's Gross Domestic Product (GDP). Glanzel et al. [Glnzel et al. 2006] and Leta et al. [Leta et al. 2006] studied Brazilian productivity and the importance of international co-authorship in Brazilian research. There are also studies of Brazilian scientific production/productivity in specific areas such psychiatry [Leta et al. 2001, Figueira et al. 2003], life sciences [Leta et al. 2005] and in limnology [Melo et al. 2006].

SCImago [SCImago 2006] is a group of Spanish universities that based in the data available from ISI, compares scientific productivity of Ibero-American universities. According to it, Brazil has three universities (USP, UNICAMP and UFRJ) among the top 10 Ibero-American universities in the area of Computer Science. One university is from Portugal and the others are from Spain.

The remainder of this paper is organized as follows: Section 2 describes the methodology used to collect the data; Section 3 describes the results regarding number of publications, distribution of the publications in different classes of journals according to their impact factor, and investments in science, research and technology for each of the country investigated in this work. Section 4 compares Brazil with different sets of countries, such as Latin American countries, Latin European countries, the BRIC (Brazil, Russia, India, and China), and other countries. Finally, Section 5 discusses the limits of this work, summarizes some important conclusions and points out possible future works.

2 Methods

All data was obtained from the ISI site [ISI 2006] in november 2006. There are 352 indexed CS journals available in the *Journal Citation Reports* (JCR) of 2005 including three Springer Lecture Notes Series (LNCS, LNAI and LNCIS) and a Proceedings of the ASIST annual meeting. For this work we did not gather data regarding the Lecture Notes Series nor the

ASIST proceedings. There is a debate within the Brazilian CS community if the Lecture Notes Series are "really journals" or mainly proceedings of conferences. Until this issue is settled by the community, we decided not to consider them journals.

We obtained all CS ISI publications from 2001 until 2005 from the following fourteen countries: *Argentina, Australia, Brazil, Chile, China, India, Israel, Italy, Mexico, Portugal, Russia, Spain, South Korea, and USA*. The data was obtained using the advanced search available at the ISI site. According to the site, one publication is considered from some country if at least one of its authors is from that country.

The specific method to retrieve the numbers in this paper is the following:

- in the JCR site, obtain all journals classified in the computer science subject area, that is, all journals in the subject areas:
  - COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE;
  - COMPUTER SCIENCE, CYBERNETICS;
  - COMPUTER SCIENCE, HARDWARE & ARCHITECTURE;
  - COMPUTER SCIENCE, INFORMATION SYSTEMS;
  - COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS;
  - COMPUTER SCIENCE, SOFTWARE ENGINEERING;
  - COMPUTER SCIENCE, THEORY & METHODS

  and remove the journals from the Lecture Notes series from the list.

- in WebofScience site, select the advanced search and enter the queries of the form:

  `PY=2001 AND CU=BRAZIL AND`
  `(SO=(`*journal1*`) OR SO=(`*journal2*`) OR...)`

  which selects the publications from Brazil for the year 2001, and where *journal1*, *journal2* and so on are journals from the list above. The query cannot contain more than 50 clauses, thus only a subset of the journals can be entered in each query.

We also provide information about the total number of indexed articles published in some given year (the same queries above without the `CU` clause). This data refereed as ALL is in the last row of Table 1.

3 Results

The results can be seen in Table 1. For each country, the table reports the total number of publications for each year from 2001 to 2005. The last row in the table (ALL), corresponds to all publications in computer science in the indexed journals in each year. The last column computes the percentage growth in publications from 2001 to 2005.

Table 2 presents the same data, but considering the percentage of publications of each country against the total number of publications. As expected, the USA, by far, publishes

Table 1: Publications for each year.

Country	2001	2002	2003	2004	2005	Growth (percentage)
Brazil	212	215	240	283	292	38%
Argentina	50	52	42	50	68	36%
Australia	579	510	562	597	681	18%
China	978	1106	1459	1756	1883	93%
Chile	35	37	44	37	48	37%
India	319	345	364	438	440	38%
Israel	398	401	414	464	454	14%
Italy	880	961	1035	1019	1072	21%
Mexico	64	82	87	95	112	75%
Portugal	100	104	146	178	175	75%
Russia	285	284	256	277	269	-6%
South Korea	574	631	724	825	797	39%
Spain	573	678	754	807	918	60%
USA	7310	7189	7798	8038	8462	16%
All	21374	21732	23674	24565	26152	22%

more than any other country in our list. But it is interesting to note that its participation on all publications has been slowly decreasing from 34,20% of the total in 2001 to 32,36% of the total in 2005.

Observing Table 1, one can separate the countries in three groups regarding the increase in the number of publications from 2001 to 2005.

The *first group* includes Australia, Israel, Italy, Russia, and USA, countries for which the increase in publications is smaller than 22% which is the total growth in the number of publications for the period. Except for Russia, this group adds to a large percentage of the total publications in CS, and each of these countries' participation in the total has been more or less constant over the years

The *second group*, with a growth around 37%, includes Argentina, Brazil, Chile, South Korea, and India. These are emerging countries, that have increased their participation as one can notice in Table 2.

The *third group* includes Mexico, Portugal, Spain and China. Each of these countries had an increase in publications of at least 60%, headed by China's 93% growth.

Russia is an exception, with growth equal to -6%, but is worth noting that its participation in all publications (Table 2) is more or less the same over the years being slightly over 1%. Nevertheless we can observe a decreasing curve.

3.1 A Categorized View of Publication Data

In this section we present a detailed view of the data considering the quality of the journals. We considered the journals sorted in non-increasing order of impact factor according to the JCR of 2005. We divided the list of journals in six groups: *A, B, C, D, E*, and *F*, each group containing 58 journals. Group A contains the 58 journals with higher impact factors. At a first approximation, higher impact factors indicate higher quality journals, but this claim must be taken with some care - different subareas in CS may have journals with very

Table 2: Percentage of all publications for each year.

Country	2001	2002	2003	2004	2005
Brazil	0,99%	0,99%	1,01%	1,15%	1,12%
Argentina	0,23%	0,24%	0,18%	0,20%	0,26%
Australia	2,71%	2,35%	2,37%	2,43%	2,60%
Chile	0,16%	0,17%	0,19%	0,15%	0,18%
China	4,58%	5,09%	6,16%	7,15%	7,20%
India	1,49%	1,59%	1,54%	1,78%	1,68%
Israel	1,86%	1,85%	1,75%	1,89%	1,74%
Italy	4,12%	4,42%	4,37%	4,15%	4,10%
Mexico	0,30%	0,38%	0,37%	0,39%	0,43%
Portugal	0,47%	0,48%	0,62%	0,72%	0,67%
Russia	1,33%	1,31%	1,08%	1,13%	1,03%
South Korea	2,69%	2,90%	3,06%	3,36%	3,05%
Spain	2,68%	3,12%	3,18%	3,29%	3,51%
USA	34,20%	33,08%	32,94%	32,72%	32,36%

different impact factors, and thus the best journals in some CS sub-area may not be in the A group (or the B group for that matter).

Table 3 reports the total number of publications for the period from 2001 to 2005, in each class, for each country. For example, Brazilian researchers had 181 publications from 2001 to 2005 in group A (the 58 journals with higher impact factors).

Table 3: Total scientific production from 2001 to 2005 by group of impact factor.

Country	A	B	C	D	E	F
Brazil	181	239	200	267	181	174
Argentina	34	69	54	53	33	19
Australia	483	651	364	575	537	319
Chile	16	34	36	50	40	25
China	1053	1105	1277	1296	1282	1169
India	274	324	356	336	339	277
Israel	653	372	436	338	210	122
Italy	898	801	857	1205	790	416
Mexico	55	87	68	87	54	89
Portugal	107	107	77	218	88	106
Russia	129	157	97	130	143	715
South Korea	378	511	521	662	544	935
Spain	611	593	557	886	647	436
USA	10274	8199	5702	6207	4557	3858
All	21624	21443	16722	21685	16233	19790

From this data, one can determine the distribution of publications of each country regarding the "quality" of journals. For example, Brazilian publications seems to be equally distributed between the six groups, whereas Russia's has the majority of its publications concentrated in group F. Table 3 shows that most of the countries have a similar distribution of its publications among the six groups, with most of the countries publishing more in the groups B, C and D.

The publications of Israel are mainly concentrated in groups A and B. The number of publications in group A represent approximately a third of all publications of the country. If one considers group B, then we can see that almost half of all of its publications are concentrated in groups A and B. USA, similarly to Israel, concentrate its publications in groups A and B.

3.2 Economic Data: GERD versus Number of Publications

The GDP (Gross Domestic Product) is a standard economic metric to measure the size of the economy of a given country. The GERD (Gross Expenditure on Research and Development) is the most commonly used metric for comparing different countries' efforts on R&D. In this section, we relate the GDP, GERD and the number of publications in Computer Science of each country.

Table 4 reports in the fifth column the total number of publications of each country in the year 2005. The second column is the GDP of 2005, in billions of dollars, normalized by purchasing-power-parity (data from the International Monetary Fund site [IMF 2006]). The third column reports the fraction of GDP destined to research and development, and the fourth column reports the amount of money, in billions of dollars, expended with research and development (data from UNESCO [UNESCO Institute for Statistics 2006, UNESCO 2006]).

The values of GERD available in [UNESCO Institute for Statistics 2006] are the most recent reported value (2000: India; 2002: Australia, Mexico; 2003: Argentina, Brazil, Chile, Italy, Korea, Portugal, Spain; 2004: China, Israel, Russia, USA). Therefore, the values of GERD in the fourth column do not refer to the same year and thus comparisons must be made with some care. The last column represents the amount money, in millions of dollars, spent for each CS publication if all the country's GERD was spent in Computer Science.

Table 4: Scientific production, GDP and GERD.

Country	GDP (Bi)	% GERD	GERD (Bi)	Publications (2005)	GERD (Mi) /Pub.
Brazil	1.576,73	0,98%	15,45	292	52,92
Argentina	533,72	0,41%	2,19	68	32,18
Australia	630,14	1,70%	10,71	681	15,73
Chile	193,21	0,61%	1,18	48	24,55
China	9.412,36	1,44%	135,54	1883	71,98
India	3.633,44	0,85%	30,88	440	70,19
Israel	158,35	4,46%	7,06	454	15,56
Italy	1.668,15	1,14%	19,02	1072	17,74
Mexico	1.072,56	0,40%	4,29	112	38,31
Portugal	203,38	0,78%	1,59	175	9,06
Russia	1.575,56	1,17%	18,43	269	68,53
South Korea	994,40	2,64%	26,25	797	32,94
Spain	1.089,10	1,11%	12,09	918	13,17
USA	12.277,58	2,68%	329,04	8462	38,88

Based on data from Table 4, we organized Table 5. The first three columns are sorted in non-increasing order of number of publications, publication growth and GERD, respectively, while the last one is sorted in non-decreasing order of the value GERD/Pub. Regarding

Table 5: Ranking based on different criteria (Number of publications, publication growth, GERD and GERD/Pub).

#	Publications	Publication Growth	GERD	GERD / Pub.
01	USA	China	USA	Portugal
02	China	Mexico	China	Spain
03	Italy	Portugal	India	Israel
04	Spain	Spain	South Korea	Australia
05	South Korea	South Korea	Italy	Italy
06	Australia	India	Russia	Chile
07	Israel	**Brazil**	**Brazil**	Argentina
08	India	Chile	Spain	South Korea
09	**Brazil**	Argentina	Australia	Mexico
10	Russia	Italy	Israel	USA
11	Portugal	Australia	Mexico	**Brazil**
12	Mexico	USA	Argentina	Russia
13	Argentina	Israel	Portugal	India
14	Chile	Russia	Chile	China

the number of publications, Brazil is the *ninth* most productive country, and considering the amount of dollars spent with research and development (GERD), Brazil is the *seventh*. Considering the last column (GERD/Pub) Brazil is the eleventh and all countries that were behind Brazil, considering GERD as the ranking, are in a better position considering the GERD/Pub as the rank. It is interesting to note that the so called BRIC countries are the less productive ones. If we consider the mean of the values GERD/Pub of the considered countries, whose value is $35,84$, we can see that these countries are very inefficient (Brazil: 52,92; Russia: 68,53; India: 70,19 and China: 71,98). However, China had the best publication growth between the years 2001 and 2005.

The ranking presented in Table 5 is just an illustrative approach. For a better estimation, one would have to use the amount of the GERD spent specifically in computer science.

4 Comparison of Brazil and other Countries

In this section we give a detailed comparison of Brazilian CS production against other countries. Figures 1 to 4 show the total number of publications from 2001 to 2005 from Brazil and other selected countries and the distribution of publications organized by the "quality" groups explained in Section 3.1.

Figure 1 compares Brazil with other Latin American countries considered in this work. Brazil is by far the country that publishes more articles in computer science. It is responsible for 57% of the published articles considering these four countries. Mexico appears in second followed by Argentina and Chile.

Figure 2 compares Brazil with other Latin European countries. Italy and Spain are the most prolific countries. It is interesting to note that Spain is getting closer to Italy considering the number of publications. One can confirm this result with the data available in Table 2, where we can see an increasing participation of Spain in the total number of

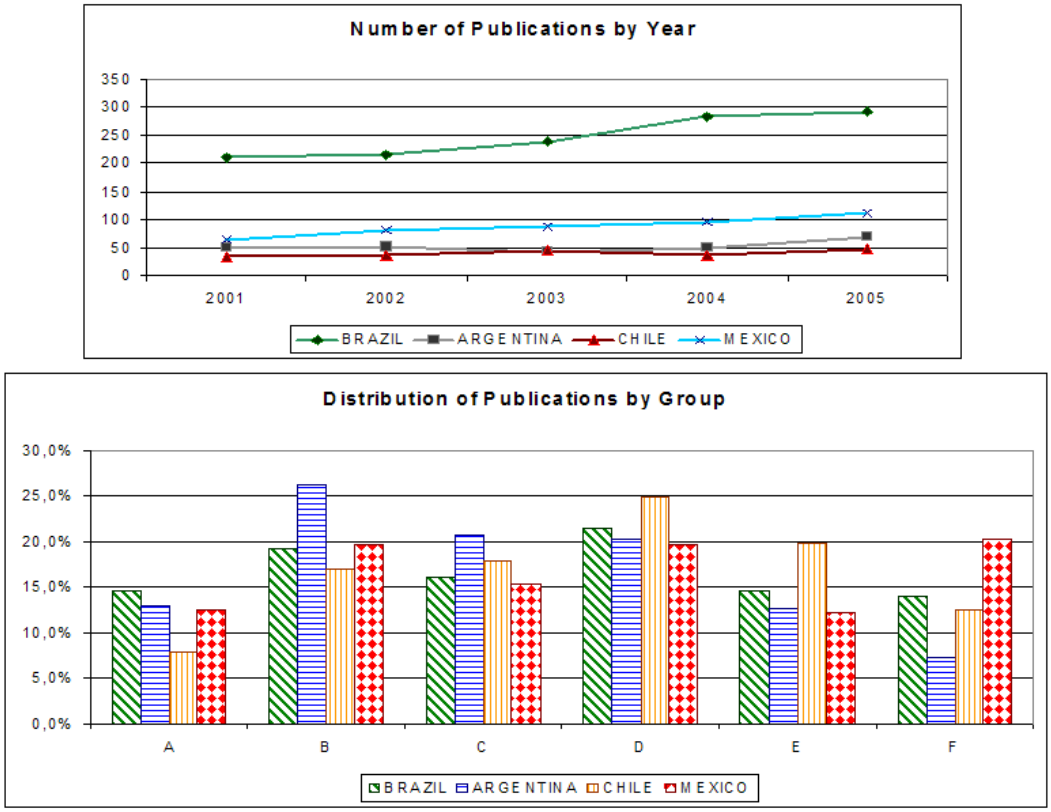

Figure 1: Latin-American countries.

publications. Italy published between three and four times more articles than Brazil, and Spain about three times more articles than Brazil.

Figure 3 compares Brazil with the other BRIC countries. China has a large production as one can see, producing four times more articles in 2005 than India, the second top country. Russia and Brazil have more or less than same number of publications, but as we commented earlier, most of the publications of Russia are in group F, considering the impact factor of journals (see the second graph of Figure 3).

Finally, Figure 4 compares Brazil with the remaining countries (except for the USA) considered in this article. South Korea appears as the leading country in this comparison, followed by Australia, Israel and at last Brazil. South Korea produced almost three times more articles than Brazil in 2005. It is worth noting in the second graph that Korea has a large number of publications in the last group of quality considering the impact factor of journals.

As we can see, Brazil has a good production considering Latin America, but in other comparisons Brazil is not well positioned. Considering the BRIC countries, China has much more publications than any other country, although Brazil has a comparable scientific production with Russia and India. Considering the Ibero-American countries Brazil is far away from Spain and Italy and considering some developed countries, Brazil is the last

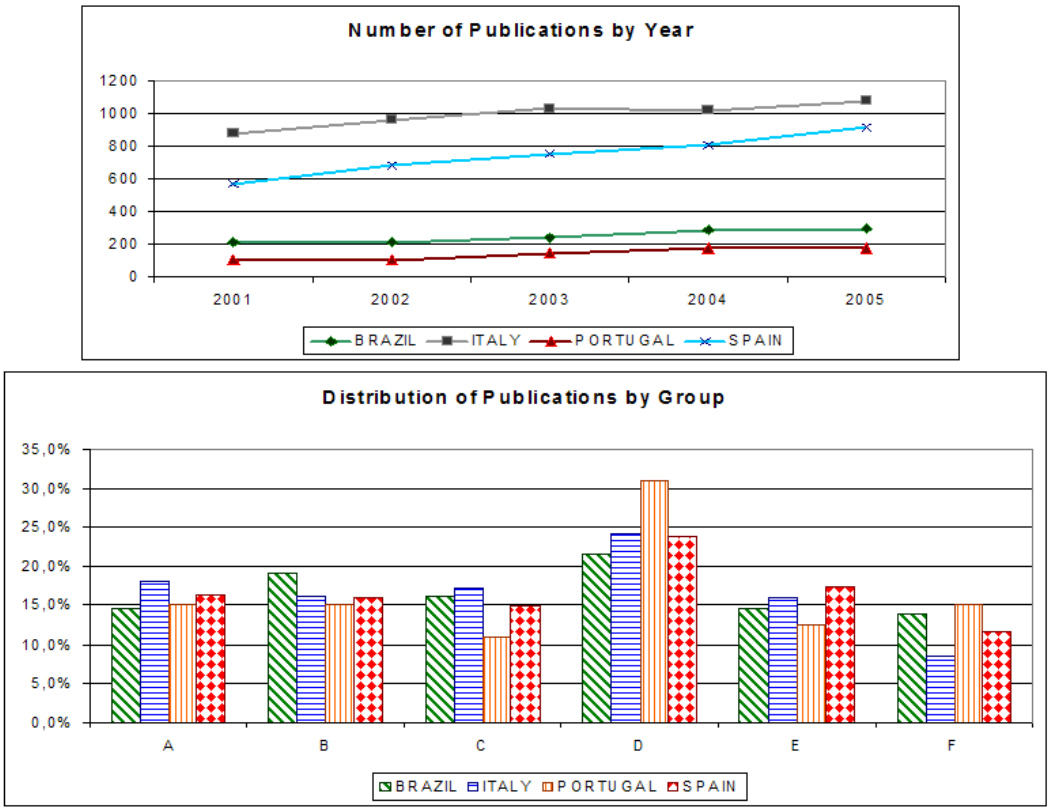

Figure 2: Latin European countries.

positioned country, although not so far from these other countries.

5 Conclusions

5.1 Overview

This work has some limitations that must be made explicit. Firstly, we used the publications indexed by ISI. But in computer science not all important and relevant publications are indexed by ISI, and most of the scientific work in computer science is not published in journals but in scientific conferences. On the other hand, it is a reasonable assumption that computer scientists in each of the countries evaluated, as a group, would distribute in the same way their work in ISI journals, non ISI journals and conferences. If this assumption is correct, then the relative values reported in this work do reflect the whole of computer science production, even if the absolute values underestimate the total production.

Second, the results in Table 4 are clearly not the correct value of investment for each publication, because not all GERD investment is spent in computer science. But if it is correct to assume that different countries assign a similar proportion of they research investment in computer science, then although the absolute values of the dollar productivity

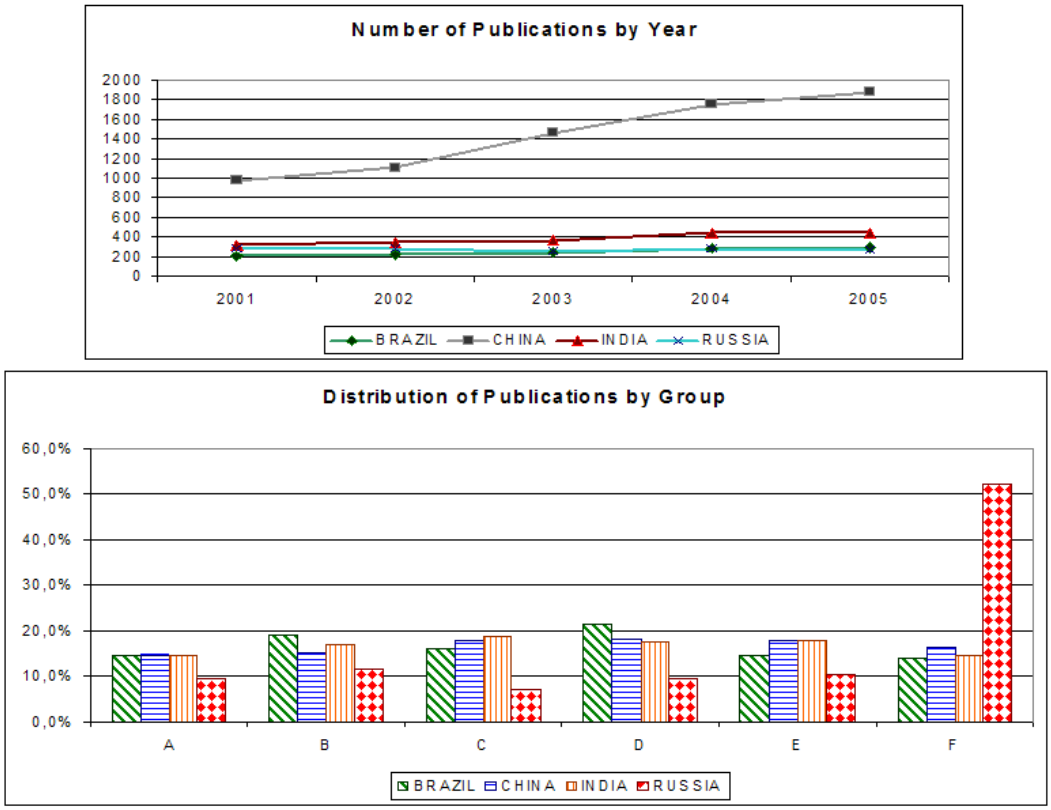

Figure 3: BRIC countries.

for each country does not reflect reality, the relative values do.

Third, as we discussed, the separation of the journals in six groups according to impact factor is only an approximation to classifying the journals according to quality.

Finally we must point out that in this paper we have not measured productivity in computer science research, but only production. To compute the productivity, one would have to divide the production figures by the total number of researchers in computer science for each of the countries in this study.

There are some interesting conclusions regarding Brazilian computer science production. Regarding Latin American countries, Brazilian production is larger than Argentina, Chile, and Mexico. The growth from 2001 to 2005 for Argentina, Brazil, and Chile are also comparable (around 37%), but much less than the Mexican growth of 75%.

Regarding other Latin countries, Brazilian production is only larger than Portugal's, but Brazilian growth of 37% is only larger than Italy's 22%, and much smaller than Portugal's growth of 75% and Spain's 60%.

The BRIC countries have all the characteristic of low dollar efficiency in their research. We cannot yet explain the result, even considering that different countries may assign different proportions of their GERD to computer science. Developing countries are likely to devote more of their research money to computer science and other technological sciences

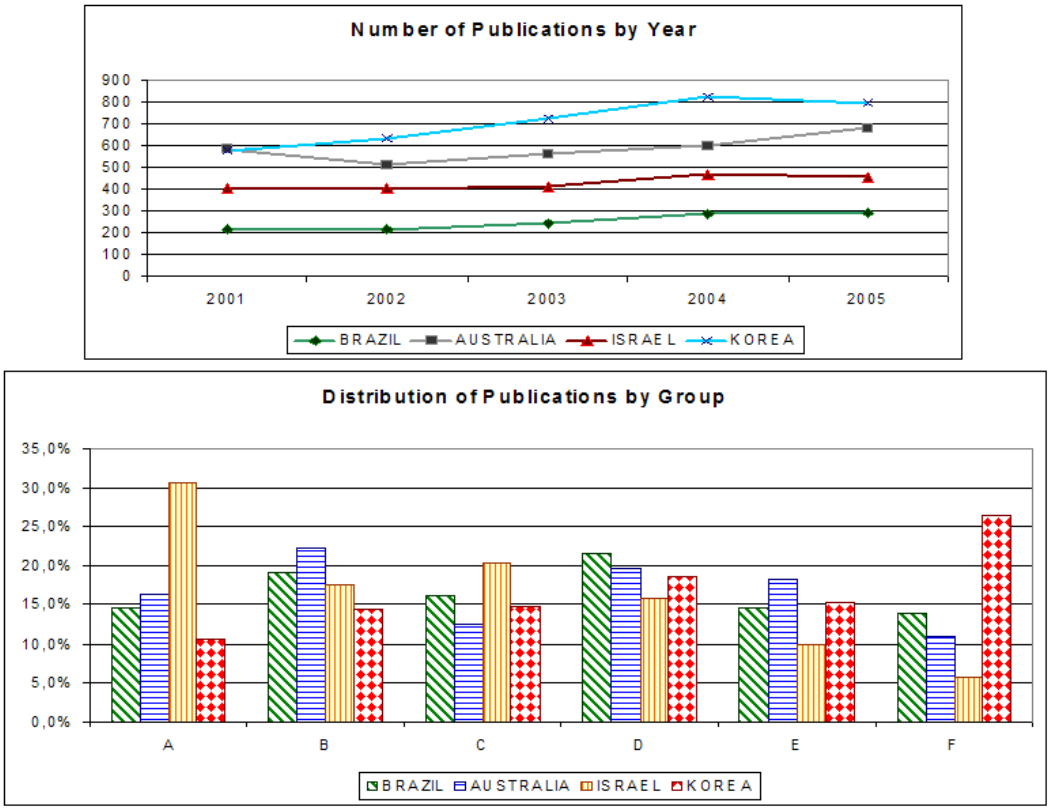

Figure 4: Brazil and other countries.

against more basic sciences, based on the belief that applied/technological research fosters innovation and increase efficiency. Thus, the inefficiency of the BRIC countries would be larger than the one shown in Table 5.

Regarding the distribution of publications in the different groups based on impact factor of the journals, the main differences seems to be concentrated in a few countries. Russia and Korea concentrate their publications in the group with lowest impact factor, whereas Israel follows the opposite direction of prioritizing higher impact factor journals.

5.2 Future Work

A possible future work is to evaluate how true are the assumptions of this work, and adapt these results in case some of the assuptions are not correct. For example, is the distribution of publications in ISI and non-ISI journals uniform across different countries? A first answer to this question may be obtained by analysing the publication patterns of a random sample of researchers from the different countries. A further understanding on how the different countries divide their CS research effort into different sub-areas is probably of value - there may be sub-areas in CS for which it is more common to publish in ISI journals, whereas for other sub-areas there may not even exist an ISI indexed journal.

But the most promising line of research is to start to evaluate the productivity of some of those countries. Australia has 2.3 times the Brazilian production in the 5 years covered by this study. That would not be a surprising result if there are 2.3 times more researchers in Australia than in Brazil. But on the other hand, if there are important differences in the productivity of Australian CS researchers in comparison with Brazilian researchers, then a deeper understanding on the way CS is carried out, the incentives and barriers the different research communities face, and other factors must be further studied. As a first approximation to the number of Brazilian computer scientists, we calculated the total number of faculty associated with departments that grant any graduate level degree in computer science in 2005. That number is around 880 people. But that figure does not include computer science researchers working in university departments that do not grant a computer science degree (for example Mathematics or Electrical Engineering), nor does it include researchers working in research centers and in industries.

To finalize, the authors believe that the Brazilian computer science research community should gain a better understanding of the processes and incentives that guide the CS research in the following countries:

- Spain
- Italy
- Mexico
- Portugal

These are all Latin countries for which the difficulties of writing research papers in English should be similar, and for which the "science culture" of competition, and evaluation of other's work should be similar. But Spain not only has a larger computer science production (three times the Brazilian production), it has grown 60% since 2001! Italy has 4 times the production of Brazil. Portugal and Mexico have shown impressive growth in the number of publications in the last 5 years. And finally, Portugal and Spain are very efficient regarding the number of publications per research investment.

References

[Albuquerque et al. 2002] Albuquerque, E. M., Simes, R., Baessa, A., Campolina, B., and Silva, L. (2002). A distribuição espacial da produção cientifica e tecnologica brasileira: uma descrição de estatisticas de produção local de patentes e artigos cientificos. *Revista Brasileira de Inovação*, 1(2):225–251. in Portuguese.

[Citeseer 2006] Citeseer (2006). Citeseer. `http://citeseer.ist.psu.edu`.

[Figueira et al. 2003] Figueira, I., Jacques, R., and Leta, J. (2003). A comparison between domestic and international publications in brazilian psychiatry. *Scientometrics*, 56(3):317–327.

[Glnzel et al. 2006] Glnzel, W., Leta, J., and Thijs, B. (2006). Science in brazil. part 1: A macro-level comparative study. *Scientometrics*, 67(1):67–86.

[IMF 2006] IMF (2006). International monetary fund. `http://www.imf.org/`.

[ISI 2006] ISI (2006). Institute for scientific information. `http://www.thomsonisi.com`.

[Leite 2006] Leite, R. C. C. (2006). O mito do atraso. Folha de São Paulo, Caderno Mais.

[Leta et al. 2006] Leta, J., Glnzel, W., and Thijs, B. (2006). Science in brazil. part 2: Sectoral and institutional research profiles. *Scientometrics*, 67(1):87–105.

[Leta et al. 2001] Leta, J., Jacques, R., Figueira, I., and Meis, L. D. (2001). Central international visibility of brazilian psychiatric publications from 1981 to 1995. *Scientometrics*, 50(1):241–254.

[Leta et al. 2005] Leta, J., Pereira, J. C. R., and Chaimovich, H. (2005). The life sciences - the relative contribution of the university of so paulo to the highest impact factor journals and to those with the largest number of articles, 1980 to 1999. *Scientometrics*, 63(3):599–616.

[Melo et al. 2006] Melo, A. S., B, L. M., and Carvalho, P. (2006). Brazilian articles in international journals on limnology. *Scientometrics*, 67(2):187–199.

[SCImago 2006] SCImago (2006). Scimago. `http://www.atlasofscience.net/atlas/scriptat/rankingf_sp.asp`.

[UNESCO 2006] UNESCO (2006). Unesco science report 2005. `http://www.unesco.org/science/psd/publications/science_report2005.shtml`.

[UNESCO Institute for Statistics 2006] UNESCO Institute for Statistics (2006). Unesco science and technology statistics. `http://stats.uis.unesco.org/ReportFolders/reportfolders.aspx`.