The Geographical Deconcentration of Scientific Activities (1987-2007)

Michel Grossetti,[*] Denis Eckert,[*] Yves Gingras,[**] Laurent Jégou,[*] Vincent Larivière[***] and Béatrice Milard[*]

[*] *rgros@servmail.univ-tlse2.fr; eckert@univ-tlse2.fr; jegou@univ-tlse2.fr; milard@univ-tlse2.fr*
Laboratoire Interdisciplinaire Solidarités, Sociétés, Territoires, CNRS and the University of Toulouse, 5 allées A. Machado, 31058 Toulouse Cedex 09.

[**] *gingras.yves@uqam.ca*
Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8 (Canada)

[***] *vincent.lariviere@umontreal.ca*
École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC. H3C 3J7 (Canada)
and
Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8 (Canada

Abstract

Traditional research on “world cities” tends to develop the idea that large, inter-connected agglomerations can better take advantage of international competition. This suggests that we should observe an increasing concentration of activities in these cities at the expense of smaller ones. Among analyses using measures based on scientific publications, certain studies support this hypothesis. Others however, show that in certain countries such as China, an opposite trend is emerging; the largest cities are undergoing a relative decline in the country's scientific activities. To go beyond this seeming contradiction, this paper provides a global analysis of all countries having papers in Thomson Reuters’ Web of Science over the period 1987-2007. The addresses present in all articles were geo-coded and then grouped into agglomerations. The result of our analysis is unambiguous: deconcentration is clearly the dominant trend both globally and within countries, despite some rare exceptions for which explanations are suggested.

Introduction

Increased geographic concentration in the largest cities is a dominant hypothesis in the current public discourse on scientific activities. The resulting key phenomenon would thus be the rise of a number of global cities, becoming dominant places of scientific production. Based on this assumption, many national scientific policies focus funding on a handful of sites deemed capable of maintaining an honorable position in global, scientific competition.

According to Matthiessen and his coauthors, the 30 largest scientific centers in the world have seen their predominance increase between 1997 and 2005: “Total growth has been 28 per cent and growth in the top 30 cities has been 34 per cent, which demonstrates a concentration process.” (Matthiessen & *al*., 2010, p. 1883). Given the potential political impact of these results, they should be carefully reviewed before scientific and urban policies are formulated too quickly. As we shall see, these ideas about concentration actually go against broader trends that, on the contrary, point to the deconcentration of both publishing and citation activities.

These publications are part of a general trend of recent analyses on the spatial organization of scientific research (Bonaccorsi & Daraio, 2005; Bornmann & Leydesdorff, 2010; Hoekman, Frenken & van Oort, 2010). While many of these works are primarily interested in specific research topics, some analyze the overall progress of scientific activity in a given country, which allows us to examine urban dynamics by neutralizing aspects linked to the country's developments. Thus, Zhou, Thijs and Glänzel (2009) demonstrated that growth in China is accompanied by a decentralization of activities between the country's different regions. Other research, based on patents (Hong, 2008) also suggests that as Chinese scientific research develops, the initial geographic concentration decreases. China's example suggests that it is better to think in terms of several spatial scales, such as the country and city levels.

Our paper aims to deconstruct the problem of decentralization by successively examining the relevant geographic levels. We begin by presenting our methodology based on a fraction-based count of publications present in the WoS, and on a multi-criteria regrouping of publication addresses in urban areas. Then, using this methodology, we will demonstrate that at the world level, publications are not more concentrated in a select group of agglomerations; rather, there is a global trend towards deconcentration.

Methods

We based our analysis on Thomson Reuters' Web of Science database, which contains the Science Citation Index Expanded. This bibliographic database lists publications (articles, notes, and reviews) from “leading, international scientific and technical journals,” that is to say, those that are globally the most cited by researchers themselves. It covers the fields of physics, chemistry, mathematics, biology (fundamental and applied), biomedical research and medicine, space and earth sciences, and engineering sciences. In 1987, it analyzed almost 500 000 publications; the total surpassed one million in 2007. Creating spatial statistics of scientific publications involves two types of choices; the first involves how publications are counted, the second the regrouping of addresses in urban areas.

Counting publications by spatial entities: the fractioning of publications

Choosing a counting unit must take into account the fact that most publications have multiple authors, grouped by affiliations, whose addresses often refer to different cities and often different countries. Each publication can thus be attributed to several geographical areas, which poses problems such as the possibility of counting the same publication many times and the challenge of regrouping spatial entities (cities within agglomerations or regions and then again within countries). Thus, if an article has a first group of authors from Leganés, a town in the suburbs of Madrid; a second group from Lugo, another suburb of Madrid; and a third in London, how should it be counted? If we count 1 for Lugo and 1 for Leganés, should we count 1 or 2 for Madrid? The first case more closely approximates the actual number of publications in which Madrid is present; but it is clear that at each spatial clustering, calculations, which are anything but simple, must be repeated. In the second case, the fact that there are two teams in Madrid is acknowledged, but the unit is no longer the article, it is the activity of the teams that signed the article, which we could define as a “scientific contribution.” To calculate Madrid's share in Spanish publications, the base would no longer be the actual number of publications but a necessarily higher sum of "contributions". If we divide Madrid's contributions by the total number of publications, the amount of shares allocated to Spanish agglomerations would be greater than 100%. Moreover, it is well known that counting articles favors large cities because of their involvement in more systematic inter-regional and international collaborations. We chose the fractional counting of publications. In our example, the Leganés team is credited with 1/3 of the publication, the same with Lugo, and with London. Madrid as a whole would be awarded 2/3 of the co-signed article. We can therefore simultaneously add up totals (as with the measure by contributions) and retain the relationship to the actual number of publications, since the sum total of all fractions would be the total number of publications worldwide.

Table 1 compares the three methods of calculating the share of world publications for 20 countries. Counting without fractioning overestimates this share, as each publication is potentially counted several times. The sum of the percentages of countries is greater than 100% (it generally tallies at 120%). The other two methods, either using the address of the first author or fractional counts, offer very similar results and adds up to 100%, which is a useful property. We consider counting fractioned publications the most rigorous method, the one that most respects scientific reality. Note, however, that the results presented in this paper were tested using all three methods. Despite minor variations, they remain remarkably stable concerning key trends.

Table 1. 1997 share of world publications in 20 countries using different methods.

COUNTRY	Counting without fractioning %	Counting by retaining only the first address %	Counting fractioned articles %
UNITED STATES	33.46	30.08	29.65
JAPAN	10.02	9.23	8.93
FRANCE	6.81	5.53	5.48
ITALY	4.31	3.57	3.36
CANADA	4.50	3.71	3.68
INDIA	2.4	2.24	2.15
SPAIN	2.84	2.4	2.30
SOUTH KOREA	1.32	1.16	1.07
AUSTRALIA	2.68	2.26	2.19
BRAZIL	1.12	0.91	0.89
NETHERLANDS	2.55	2.01	1.99
TURKEY	0.57	0.51	0.48
TAIWAN	1.22	1.12	0.98
POLAND	1.24	0.95	0.94
SWEDEN	2.06	1.63	1.59
SWITZERLAND	1.86	1.35	1.31
BELGIUM	1.35	1	0.97
IRAN	0.1	0.09	0.09
ISRAEL	1.23	0.99	0.95
GREECE	0.59	0.48	0.45

Building agglomerations: combining criteria

The scientific “localities” we determined constitute the most accurate publishing points that our geographic coding method allows us to localize. This is the smallest unit of tracking possible using the WoS “address” field. However, the level of precision obtained varies considerably between cities and countries. This is due to the fact that generally, the mailing address contained in this field corresponds to the municipality where the research organization is located. This spatial entity's size can vary considerably from one country to another depending on the degree of administrative fragmentation. How can publishing entities be regrouped within urban entities based on uniform criteria for every country in the world? We have observed the significant difficulties that have faced researchers who have approached this task (Matthiessen & *al*., 2010), even when they limited themselves to several dozen scientific centers. Our aim was to produce universal criteria, and not divisions corresponding to a juxtaposition of national criteria (for example, using SMSAs for the United States, urban areas in France, etc.., and then comparing the results). After confronting global data sets of different sorts, all of which are open access - on land occupation for some, such as ESA Iona GlobCover or Global UrbanExtent (for the last: see Schneider & *al*., 2009, and 2010), and on population densities for others (Global Population of the World), we concluded that the best data to delineate all the cities of the world were those related to population density (highly fine-tuned raster data). Indeed, data from land artificialization are not sufficiently discriminating. Such is the case in continuously constructed coastal areas (often designed for tourists), which does not necessarily correspond to year-round, continuous human occupation, nor necessarily dense inhabitation, and even less likely to harbor areas of scientific activity.

Using data on population density is more convincing. The best approach was to use an indicator to spatially determine strong variations in density, in order to help us produce a usable delimitation of urban areas. Among the *Local Indicators of Spatial Association*, we chose the local Moran's I values that identify significantly dense nuclei (Anselin, 1995). Processing density data leads to a homogeneous criterion for major conurbations. For smaller urban entities, a simpler criterion of distance (a threshold of 40 km between two localities producing publications) allowed us to define “small” scientific cities.

Results

Deconcentration of publications in all the cities in the world

A simple way to evaluate concentration is to calculate the proportion of world publications produced by a given number of productive cities, say first 10, first 20, first 100, etc (Table 2). The results are unambiguous: the trend towards decentralization is general between 1987 and 2007.

The proportion of world publications by the 25 cities ranked among the top 30 in both 1997 and 2007 regressed by 3.72%. The share of the 5 other agglomerations regressed by 0.67% during the same period. Unsurprisingly, the share of newcomers increased by 1.13%. However, more interestingly, it is the other cities that together increased by 3.26%. Note that though the analysis is based on growth rate, our results do not concur with those of Matthiessen & *al*. (2010): the 30 major cities in 2007 increased their publications by 28% between 1997 and 2007, while the world increase totaled 40%. This difference is probably due to the method of calculation; fractioning avoids the overestimation produced by calculating per publication.

Table 2. Changes in the concentration of world publications in the most important cities

Years Agglomerations	Share of total publications 1986-1988 (Moving average)	Share of total publications 1996-1998 (Moving average)	Share of total publications 2006-2008 (Moving average)
First 10	21,0%	16,8%	13,5%
First 20	28,7%	24,5%	21,4%
First 30	34,1%	30,0%	26,7%
First 50	42,3%	39,0%	35,1%
First 100	56 ,3%	52,5%	49,0%
First 200	72,5%	69,6%	65,7%
First 300	79,5%	78,3%	75,1%

To assess the extent to which the evolution of cities is linked to that of their country, we carried out an ANOVA on the agglomerations of the 14 most productive scientific countries.[1] This analysis shows that the rate of increase in the number of publications per country explains 72% of differences in publication rates between cities[2]. It is therefore necessary to proceed in stages and to examine the changing balance between countries before analyzing their cities.

At the country level, scientific activity is deconcentrating

Table 3 shows our calculations for the 30 countries with the most publications in 2007; these calculations fully converged with existing analyses. They show that deconcentration is clear at the country level. In 1987, three countries (USA, UK, and Germany) alone produced 50% of publications. Nine countries made up 75% of publications and with 20, the total reached 90%. In 1997, four countries (USA, Japan, Germany, and the United Kingdom) accounted for 50% of publications. Eleven produced 75% of the world's production and 23 countries made up 90%. In 2007, the numbers are 5 countries (USA, China, Japan, Germany, and the UK) for 50%, 14 for 75%, and 26 for 90%.

[1] United States, China, South Korea, Japan, France, Canada, Germany, United Kingdom, Brazil, India, Spain, Italy, Russia, Australia. These countries account for 72.7% of the total number of world publications.
[2] R Squared = 0,728 (Adjusted R Squared = 0,715).

Table 3. Evolution between 1987 and 2007 of the 30 most productive countries in world publication.

Country	Country/ World share in 1987 *	Country / World share in 1997 *	Country/ World share in 2007 *	1987 Rank	1997 Rank	2007 Rank
UNITED STATES	34.3	29.8	24.5	1	1	1
CHINA	0.9	2.4	8.6	18	12	2
JAPAN	7.3	9.0	7.8	4	2	3
GERMANY	7.4	7.4	5.9	3	3	4
UNITED KINGDOM	7.8	7.4	5.4	2	4	5
FRANCE	5.3.	5.5	4.3	5	5	6
ITALY	2.5	3.4	3.5	9	8	7
CANADA	4.3	3.7	3.4	7	6	8
INDIA	2.6	2.2	3.0	8	11	9
SPAIN	1.3	2.3	2.7	13	9	10
SOUTH KOREA	0.1	1.1	2.7	42	16	11
RUSSIA	5.1	3.5	2.2	6	7	12
AUSTRALIA	2.1	2.2	2.1	10	10	13
BRAZIL	0.5	0.9	2.1	27	21	14
NETHERLANDS	1.8	2.0	1.7	11	13	15
TURKEY	0.1	0.5	1.6	44	26	16
TAIWAN	0.3	1.0	1.6	33	18	17
POLAND	1.0	0.9	1.4	17	20	18
SWEDEN	1.6	1.6	1.2	12	14	19
SWITZERLAND	1.2	1.3	1.2	14	15	20
BELGIUM	0.9	1.0	0.9	19	17	21
IRAN	0.0	0.1	0.8	73	51	22
ISRAEL	1.0	0.9	0.8	16	19	23
GREECE	0.3	0.5	0.7	31	28	24.
AUSTRIA	0.6	0.7	0.6	23	24	25
DENMARK	0.8	0.8	0.6	20	22	26
FINLAND	0.6	0.7	0.6	21	23	27
MEXICO	0.2	0,4	0.6	35	32	28
CZECH REPUBLIC	0.5	0,4	0.5	24	29	29
SINGAPORE	0.1	0.3	0.5	45	35	30
World TOTAL	481044	654550	918564			

* Fractional counting of publications, with 3-year moving average.

At the country level, scientific activity is deconcentrating

There is a clear shift in the balance of scientific publications between countries, but it could very well be accompanied by an increased concentration in some major scientific cities, as suggested by Matthiessen & *al*. (2010). To test this hypothesis, we will now examine how the share of publications of large scientific cities has changed with respect to the total publications of their respective countries (Table 4).

Table 4. Changes in the share of publications of the world's major agglomerations in terms of their respective national production between 1987 and 2007.

Agglomeration	City / country share 1987	City / country share 1997	City / country share 2007	Changes in city / country shares 07-97
TOKYO	34.1	32.5	32.5	0.0
BEIJING	34.4	26.5	21.1	-5.4
PARIS	45.2	39.0	35.7	-3.3
NEW YORK	7.2	6.5	6.0	-0.5
SEOUL	82.6	48.5	54.2	5.7
BOSTON	5.2	5.5	5.7	0.2
KYOTO	21.7	20.7	19.8	-0.9
LONDON	27.1	23.5	22.2	-1.3
BERKELEY	5.4	5.0	4.6	-0.5
LOS ANGELES	4.7	4.2	4.4	0.2
WASHINGTON	5.5	5.0	4.4	-0.6
SHANGHAI	18.4	11.1	10.8	-0.2
MOSCOW	57.6	47.5	42.8	-4.8
TAIPEI	51.4	51.1	44.5	-6.5
TORONTO	19,6	19.5	20.6	1.1
CHICAGO	3.0	2.9	2.8	-0.1
MADRID	31.1	26.8	23.7	-3.2
PHILADELPHIA	2.9	2.9	2.6	-0.3
BERLIN	10.0	9.9	10.2	0.3
DURHAM RESEARCH TRIANGLE	2.0	2.1	2.3	0.2
SYDNEY	24.1	26.0	25.4	-0.6
ROME	15.0	15.7	15.2	0.5
COLOGNE	10.1	9.8	9.0	-0.8
SINGAPORE	100.0	100.0	100.0	0.0
BALTIMORE	01/07/12	1.9	2.1	0.2
BARCELONA	19.8	19.4	18.8	-0.6
MUNICH	9.7	09/01/12	8.8	-0.3
MELBOURNE	24.6	22.8	23.9	1.1
MONTREAL	13.8	16.3	15.1	-1.2
SAN DIEGO	1.6	2.0	2.0	0.0

Of the 31 most important cities in 2007, Singapore must be excluded because deconcentration is not possible in the city-state. Of the remaining 30, 20 saw their share decline, two were stable, and 8 increased their ranking.

Discussion and conclusions

Contrary to the prevalent belief in many debates and decisions on scientific policy, the overall trend is not towards a concentration of scientific activities in "global cities." Rather, we are observing what is primarily a loss of hegemony of the usual central countries (USA, UK, Germany, France), accompanied by particularly strong growth in Asian countries (China, South Korea, Taiwan), and more broadly many "emerging" countries. This development contributes to diversifying production sites of scientific papers. Furthermore, we have observed within many countries (Russia, France, Spain, Great Britain, China, etc.), regardless of the overall development of the national scientific production, a trend towards deconcentration with increased production in “secondary” cities. Those countries that experienced the earliest deconcentrations (USA, Canada, Germany, for example) have a rather stable scientific geography. Countries currently concentrating their scientific production are rare and explanations can be found for each in the changes in their demography, economy, and scientific organization.

This paper has focused only on the production of publications. The next step would be to analyze the impact of these publications by examining how they are cited. Recent studies show that there is no link between the location of a highly cited researcher and the citation rate of the city in which he is located (Bornmann & Leydesdorff, 2011). Moreover, a study on France (Grossetti and Milard, 2011) shows that even though citation rates of publications written by researchers in cities outside Paris remain ever so slightly lower than those for Paris publications, the former tend to follow closely the latter. Finally, it has been shown that among all publications, there is a clear deconcentration of citations over time at the world level (Larivière & *al*., 2010). The same deconcentration has also been shown for China using the Chinese Citation Index (Yang & *al*., 2010).

All these results are consistent with the general conclusion that scientific activities are more widespread geographically as well as more visible than ever and that there is no real trend towards concentrating activities in so-called “world-cities,” despite widespread belief. As the planet's overall economic and social structures are changing, scientific activities are spreading across a wider geographic area. Our data and analysis show that the system of cities of scientific research is evolving, leading to newer and more numerous and dynamic nodes of scientific production throughout the globe.

References

Anselin, L. (1995). Local Indicators of Spatial Association—LISA, *Geographical Analysis*, 27 (2), 93-115.

Bonaccorsi, A. & Daraio, C. (2005). Exploring Size and Agglomeration Effects on Public Research Productivity, *Scientometrics*, 63 (1), 87-120.

Bornmann, L. & Leydesdorff, L. (2011). Which Cities Produce More Excellent Papers than Can Be Expected? A New Mapping Approach—Using Google Maps—based on Statistical Significance Testing. *Journal of the American Society for Information Science and Technology*, 62 (10), 1954-1962.

Grossetti, M. & Milard, B. (2011). La concentration spatiale des activités scientifiques en question. *Proceedings of the Third International RESUP Conference “Reforming Higher Education and Research.*” Available at http://www.resup.u-bordeaux2.fr/manifestations/colloque%20international%20Paris%202011/Colloque%20RESUP%202011%20Vol%202.pdf

Hoekman, J., Frenken, K. & Tijssen, R.J. (2010). Research Collaboration at a Distance: Changing Spatial Patterns of Scientific Collaboration within Europe. *Research Policy*, 39 (5), 662-673.

Hong, W. (2008). Decline of the Center: The Decentralizing Process of Knowledge Transfer of Chinese Universities from 1985 to 2004. *Research Policy*, 37 (4), 580-595.

Larivière, V., Gingras, Y. & Archambault, E. (2009). The Decline in the Concentration of Citations, 1900-2007. *Journal of the American Society for Information Science and Technology*, 60 (4), 858-862.

Matthiessen, C.W., Winkel, A.S. & Find, S. (2010). World Cities of Scientific Knowledge: Systems, Networks, and Potential Dynamics. *Urban Studies*, 47 (9), 1879-1897.

Schneider, A., Friedl, M. & Potere, D. (2009). A New Map of Global Urban Extent from MODIS Data. *Environmental Research Letters*, 4 (4), article 044003.

Schneider, A., Friedl, M. & Potere, D. (2010). Mapping Global Urban Areas Using MODIS 500-m Data: New Methods and datasets based on 'Urban Ecoregions'. *Remote Sensing of Environment*, 114 (8), 1733-1746.

Yang, S., Ma, F., Song, Y. & Qiu, J. (2010). A Longitudinal Analysis of Citation Distribution Breadth for Chinese Scholars. *Scientometrics*, 85 (3), 755-765.

Zhou, P., Thijs, B. & Glänzel, W. (2009). Regional Analysis on Chinese Scientific Output. *Scientometrics*, 81 (3), 839-857.