

The Geography of Chinese Science

David Emanuel Andersson^{a,*}, Søren Find^b, Saileshsingh Gunessee^a, Christian Wichmann
Matthiessen^c

^a *Nottingham University Business School China, 199 Taikang East Road, Ningbo 315100, Zhejiang, China*

^b *Technical University of Denmark, Anker Engelunds Vej 1, Kgs. Lyngby 2800, Denmark*

^c *Department of Geosciences and Resource Management, University of Copenhagen, Øster Voldgade 10, København K 3460, Denmark*

ABSTRACT

Chinese scientific output has increased dramatically in recent years, but its internal spatial structure has received scant attention. Estimated gravity models of intercity scientific co-authorships show that there are two types of spatial political bias in China, apart from the expected mass and distance effects. Intercity co-authorships involving Beijing are more common than Beijing's output volume and location would imply, and this Beijing bias is increasing over time. The second type of spatial political bias is greater intra-provincial collaboration than is accounted for by size and distance. The geography of Chinese science is thus not only monocentric as regards overall scientific output, but also exhibits unusually hierarchical collaboration patterns. Unlike in Europe and North America, national and regional capitals are becoming ever more important as scientific coordination centers.

Keywords: scientific collaboration, network, China, spatial political bias

* Corresponding author. Tel.: +86 574 8818 0928; Fax: +86 574 8818 0125.

E-mail: david.andersson@nottingham.edu.cn (D.E. Andersson); sf@vitis.dtu.dk (S. Find), saileshsingh.gunessee@nottingham.edu.cn (S. Gunessee), cwm@geogr.ku.dk (C.W. Matthiessen).

1. Introduction

The recent rise of Chinese science has attracted considerable attention (Nature, 2013; Royal Society, 2011; Zhou and Leydesdorff, 2006). While China's annual economic growth amounted to about ten percent in the first decade of the twenty-first century, its scientific growth averaged 18 percent per year between 1996 and 2008 (Royal Society, 2011, p. 20). China's scientific growth is thus even more dramatic than its economic growth.

While this rise has resulted in a number of scholarly articles (Liefner and Hennemann, 2011; Zhou and Leydesdorff, 2006; Zhou et al, 2009), we still know little about the *internal spatial structure* of Chinese science. In particular, previous studies do not address the role that geography has played in scientific interaction *among cities* in China.¹ Most such studies are descriptive and take a national or, at best, a provincial perspective².

The aim of this paper is to understand the spatial structure of Chinese science, using an econometric application of the well-known gravity model of interacting nodes. It is widely known that Beijing has greater science output than can be accounted for by the size of its population or economy, even when compared with Chinese cities of otherwise similar size and sophistication such as Shanghai or Shenzhen. What is less well known is whether size and distance are the only important factors that explain the *interactivity* of Chinese science, which is what one would expect in market-driven spatial structures, or whether political prioritization of national and/or regional capitals introduces a *spatial political bias*. Statistical testing of potential spatial bias is our primary objective. A second question, which is conditional on the

¹ It is scientists within cities that collaborate and not cities per se, but cities function as nodes that host scientists. In the intercity network literature this is referred to as a two-mode network with links between two different entities (scientists and cities) which is transformed into a one-mode network involving only one mode such as cities (see Liu et al., 2013).

² This relates to two aspects: the geographical scope (e.g. continent or nation) and the spatial unit (e.g. country or city region). Most studies focus on other scope-unit combinations than a nation and its city regions. Some studies overlap with this study by focusing on either cities *or* on a single nation. For example, Matthiessen et al. (2002; 2010) provide a global network analysis of cities whereas Havemann et al. (2006) use trend analysis to study collaboration among German immunological institutes. The closest analogy to this study is Ponds et al. (2007), which analyzes collaboration between Dutch NUTS3 regions. Frenken et al. (2009) provide a literature survey and observe that "studies on collaboration patterns among regions or cities are rare."

existence of spatial political bias, is whether this bias has changed over time. Studies of European interaction show that national border effects have been decreasing in recent years.

A supplementary question is to compare the efficacy of different measures of mass. The most common measure is the number of publications, which is easily observed but may be affected by endogeneity problems. Therefore we also include exogenous approximations of market or political input volumes such as city GDP and the number of national research universities in a city.

The paper proceeds as follows. The next section summarizes the recent growth of Chinese science, while section 3 introduces the application of gravity models and other approaches to scientific collaboration. Section 4 presents methods and a descriptive overview of the dataset, while section 5 presents the econometric results. The penultimate section is an exploratory extension that attempts to “read between the lines” of the econometric estimates. Section 7 concludes.

2. The rise of Chinese science

Table 1 shows the world’s top 12 science city regions in three distinct periods. While Europe and North America dominate global rankings of science cities, these rankings are not entirely static. The two most conspicuous changes after 2000 have been the rise of three East Asian cities—Beijing, Shanghai, and Seoul—and the decline of Moscow. In 1996-98 there were no Chinese cities among the top-12 city regions, while by 2010 Beijing and Shanghai had joined this group, with Beijing ranked first.

TABLE 1 ABOUT HERE

TABLE 2 ABOUT HERE

Table 2 lists 31 leading Chinese science cities in 2008-2010 and their growth rates from 1996-1998 to 2008-2010. Beijing and Shanghai are not the only Chinese centers. Among the world's 75 largest science cities in 2008-2010, nine were in mainland China. These nine Chinese cities exhibited the nine highest SCI growth rates among the 75 cities between 1996 and 2010.

All 31 Chinese cities had high growth rates, not only the top nine. If we compare their growth with the top 66 cities outside China, only two cities (Lanzhou and Changchun) had lower growth rates than the fastest-growing non-Chinese city, which was Seoul. Even so, these two lagging Chinese cities expanded faster than São Paulo, which ranked second among non-Chinese cities. The stylized facts thus suggest that there is an ongoing process whereby new science cities are continually emerging.

3. Theoretical background

When investigating scientific collaboration in space, the gravity model is appropriate as it is the key model for analyzing spatial interaction. It has enjoyed considerable success in various empirical applications and rests on sound theoretical micro-foundations. In what follows we provide an overview of the gravity model of scientific knowledge flows, as well as a brief outline of the most popular alternative, which is the network approach.

3.1. The gravity approach

Theory. Building on earlier literature³, Beckmann (1993; 1994; 1999) provides a gravity model of scientific knowledge flows. His probabilistic micro-level theory assumes that two rational agents (scientists) supply inputs into joint scientific production. The production function exhibits positive marginal products, diminishing returns to substitution and constant returns to scale. The agents maximize collaborative scientific production net of costs when choosing their labor effort; costs include the time- and distance-dependent costs of achieving effective interaction. Through a series of substitutions and extensions, Beckmann obtains a gravity equation with scientific knowledge flows between two locations, i and j :

$$C_{ij} = N_i \cdot N_j \cdot e^{-d_{ij}} \quad (1)$$

where N is the number of researchers in i and j and d is the distance between the two locations.

Andersson and Persson (1993) and Andersson and Andersson (2006, pp. 215-16) offer a deterministic alternative to Beckmann's model. The main assumption is that the value of scientific activities is dependent on ideas and techniques generated in other cities than the scientist's home region. This is represented by P -functions, which depend on the inflow of ideas coming from other nodes (cities) in the science network. Such ideas can come in the form of information (I) or knowledge (K). While information is easily transmitted by communication modes such as the Internet, knowledge flows more often require face-to-face contacts.

Further assumptions include that the value (P) and production (Q) functions are differentiable and concave with respect to inputs. To determine the optimal flows of local and interactive inputs, we maximize net value, G_i :

³ Isard (1960) documents early gravity applications in the social sciences, which served as precursors of later models and applications. Wilson (1970) as well as Sen and Smith (1994) provide the foundation for modeling probabilistic spatial interaction behavior, such as Beckmann (1993; 1994; 1999). Such models have been applied to the study of traffic flows, interregional trade and migration flows.

$$G_i = P_i(I, K) * Q(L_i, K_i, I_i) - \sum_{j \neq i} \tau_{ji} I_{ji} - \sum_{j \neq i} \omega_{ji} K_{ji} - \omega_i K_i - \tau_i I_i - w_i L_i \quad (2)$$

where $I = \{I_{11}, \dots, I_{ji}, \dots, I_{nn}\}$ is a vector of information flows from node (city) j to node (city) i ; $K = \{K_{11}, \dots, K_{ji}, \dots, K_{nn}\}$ is a vector of knowledge flows from j to i ; L_i is labor in city i ; P_i is quality as measured by publication or citation score of city i ; Q_i = science output in city i ; τ is price of information, including unit transaction and communication cost; ω = price of knowledge, including unit transaction and communication cost; w = price (wage) of labor.

Maximization yields the marginal interactivity conditions:

$$\begin{cases} \frac{\partial P_i}{\partial I_{ij}} Q_i - \tau_{ji} = 0; & j = 1, \dots, n; i = 1, \dots, n; i \neq j \\ \frac{\partial P_i}{\partial K_{ij}} Q_i - \omega_{ji} = 0 & j = 1, \dots, n; i = 1, \dots, n; i \neq j \end{cases} \quad (3)$$

These marginal interactivity conditions exhibit *economies of scale* in long-distance interaction, implying that the larger a node's scientific agglomeration is, the more globalized it will be in terms of the size and spatial scope of its interaction pattern. Working with a simplified form for $\partial P / \partial K$, the interactions K_{ij} follow a gravity function:

$$K_{ij} = Q_i^{\lambda_1} Q_j^{\lambda_2} f(d_{ij}) \quad (4)$$

where Q is scientific output (mass) in locations i and j , while λ_1 and λ_2 are parameters that may be estimated empirically; $f(d_{ij})$ is a distance friction function that reflects the impact of distance on space-bridging costs.

The gravity model thus implies that scientific collaboration between two locations increases proportionally with the product of the mass variables but declines with distance. Although both models arrive at gravity formulations that conform to feasible empirical analyses, only Beckmann's model uses inputs as mass variables. Thus, Beckmann's model implies the use of input variables such as the number of scientists rather than scientists' output. As far as we know, all published gravity-type estimations of scientific interactivity use output volumes, and thus conform to the deterministic model (Andersson and Persson, 1993; Andersson and Andersson, 2006).

Comparable studies. The gravity approach suggests that one should account for geographic proximity, since distance is often the best explanation of interaction quantities involving cities. Proximity facilitates face-to-face communication, which in turn facilitates the transmission of tacit knowledge as well as serendipitous discoveries, both of which stimulate the production of creative outputs such as scientific publications (Andersson and Persson, 1993). Moreover, face-to-face interaction may increase interpersonal trust, making it easier for people to collaborate (Ponds et al., 2007).

Gravity models allow researchers to study how the effects of various factors *change over time*. Most studies that have examined distance effects in science have however adopted a static framework, such as analyzing collaboration at one point in time or using pooled observations from several years (Hansen, 2013). There are a few exceptions, but Hoekman et al. (2010) is the only gravity-type analysis that explicitly addresses change. Other things being equal, we would expect a decreasing effect of distance over time, since space-bridging costs have been falling as a result of technological progress and investments in transport and communication infrastructures.

There are only a handful of studies that analyze scientific collaboration with the help of gravity models (Acosta et al, 2011; Andersson and Persson, 1993; Hoekman et al, 2010; Hoekman et al, 2009; Ponds et al, 2007; Scherngell and Hu, 2011)⁴. The distance effect is uniformly negative but the estimated magnitudes depend on the chosen measurement technique⁵. The distance effects in studies that resemble the present study range from -.23 to -.70. Hoekman et al. (2010) show that in Europe, the distance effect increased from 2000 to 2007. They explain this increase as reflecting the emergence of collaboration involving scientists in peripheral European regions, which will tend to cause a greater measured distance effect than if collaboration had remained confined to regions closer to the European core. Given China's size, we expect distance to be as important as it is in Europe. Still, the relevant question is not whether distance matters but *how much* it matters; how does it compare to distance frictions elsewhere and how does it evolve over time?

Spatial political bias. It is well known that a substantial share of Chinese science funding is attributable to governmental decision-makers. In spite of recent globalizing tendencies, Jonkers (2010) contends that Chinese science is still a top-down system with little bottom-up scientist-driven research. Moreover, leading universities are with few exceptions located in Beijing or in provincial capital cities. It is thus likely that scientific collaboration patterns reflect political resource allocation decisions⁶. Jonkers (2010, p. 36) claims that this political influence implies

⁴ Only six studies of Chinese scientific collaboration address collaboration between spatially delimited areas and only one (Scherngell and Hu, 2011) uses a gravity estimation. Our gravity-type model differs from Scherngell and Hu in three important ways. First, we use the global SCI database rather than a domestic Chinese database. This makes it possible to compare Chinese output and interactivity with cities outside China. Second, our spatial unit of analysis is a functional urban region rather than a province. The only provinces that approximate functional urban regions are Beijing, Shanghai and Tianjin. Third, they analyze scientific collaboration for one single point in time (i.e. 2007), which precludes panel techniques and analysis of dynamics.

⁵ Some studies use alternative methods when examining the influence of distance (Katz, 1994; Liang and Zhu, 2002). Overall, comparable studies support the hypothesis that geographical proximity matters.

⁶ Our explanation is consistent with a more general explanation of a "Confucian hierarchical education system" which is defined as being molded by the national government, with a national entrance examination that supports a hierarchical university system. Top universities in the national system have three advantages: the best students, the most research funding and the most attractive terms and conditions for faculty. These advantages mutually

a preference for funding large-scale collaborative projects in fields that policy-makers deem important. He also claims (2010, pp. 148-49) that there is a network of Chinese scientists who jointly decide on the allocation of research funds.

One would thus expect political priorities to matter more than economic factors in science-related location choices. One illustration is the priority given to national or provincial capitals as the preferred locations for national universities. The booming centers for foreign direct investment and exports—primarily Ningbo, Shenzhen and Suzhou—account for tiny shares of China’s scientific publications, in spite of their high growth rates and per-capita incomes (see Table 2).

A hypothesis that conforms to the notion of spatial political bias is that we should expect the funding of scientific activities—including collaborative research projects—to prioritize Beijing as well as provincial capitals. The implication is an expected over-representation of linkages involving Beijing and provincial capitals.

Spatial political bias: same-province effect. The distance effect is just one facet of what a gravity model can measure. Though the theoretical gravity models do not include political biases such as capital-city over-representation, it is common to extend the basic model to account for such effects. Empirical studies of international science networks provide estimates of how international border crossings reduce interactivity between cities (Hoekman et al, 2010; Okubo and Zitt, 2004)⁷. While linguistic or cultural barriers are less important in China than in Europe, political barriers between provinces cannot be ruled out in light of the practice of

reinforce the dominance of established top universities such as Peking University, Tsinghua and Fudan in China. Seen from this perspective, the mainland Chinese system resembles the research and education systems of Japan, South Korea and Taiwan (see Marginson, 2011).

⁷ There is a literature that looks at how national political or linguistic biases favor national over international collaboration. For example, Okubo and Zitt (2004) show that French border regions cooperate little with regions on the other side of the border, with only the Paris region exhibiting strong international linkages in science.

Chinese science policy, which includes a division of labor between different levels of the spatial political hierarchy, such as nation, province and city.

Indeed, Scherngell and Hu (2011) contend that regional protectionism as a manifestation of spatial political bias is pervasive in both science and industry. Others have argued that provincial governments are inward-oriented in their science policies, protecting local institutes and universities with the aim of maximizing intra-provincial benefits (Chen and Wang, 2003; Yoon, 2011). However, assertions of provincial protectionism in science have so far relied on anecdotal evidence rather than econometric estimates.

In their study of Europe, Hoekman et al. (2010) estimated not only the effects of national barriers, but also found that such barriers became less important between 2000 and 2010. There is thus some empirical support for the idea that European science is becoming more integrated. The Chinese analogy to Europe's national border effects is the "same-province effect," which refers to the hypothesis that provincial funding organizations prefer to keep their funds within the same province.

3.2. *Other theoretical approaches*

The most common non-gravity approaches for analyzing interregional scientific collaboration are various adaptations of network theory⁸. The network approach provides a different way of examining scientific co-authorships, with the aim of identifying sub-networks or clusters of cities with stronger-than-average inter-linkages⁹.

⁸ The "scientific collaboration in space" literature is mostly data-driven and lacks a clear theoretical framework. Most studies offer descriptive analyses of collaboration, using trend analysis, matrix-based approaches or indices to measure various aspects of scientific collaboration, including geographical proximity as one such aspect (see Havemann et al, 2006; Katz, 1994; Liang and Zhu, 2002).

⁹ Most network studies do not use the city region as the analyzed spatial unit due to limited data availability (Frenken et al., 2009; White, 2011). Studies of interregional networks tend to infer connectivity from metrics of network centralization and clustering (Oner et al., 2010). Most network studies lack theoretical micro-foundations. Liefner and Hennemann (2011) is probably the best attempt to provide a theoretical framework to connect network theory to regional spatial phenomena.

An approach that focuses on “the world network of science cities” has been the most influential alternative to the gravity model in generating empirical studies. This approach takes its cue from Taylor (2004) and his notion of “world city networks.” Matthiessen et al. (2002; 2010; 2011) uses Science Citation Index (SCI) data to classify cities as belonging to different hierarchical “levels” and “bands” on the basis of their volume and interactivity. These attempts make use of a direct analogy of Taylor’s (2004) classification, where “world city” status depends on the locations of headquarters and offices of multinational corporations.

While existing network approaches to science cities provide a summary of the relative importance of different cities, it remains the case that Taylor and his followers do not provide any micro-analytic theoretical foundation, in contrast with the gravity model¹⁰. There are thus strong theoretical reasons for giving priority to the gravity model as the preferred starting point for quantitative analyses of inter-city links.

4. Data and Methods

4.1. Data

Scientific co-authorships remain the main form of scientific collaborative output (Ponds et al, 2007). SCI-indexed co-authorships comprise all published articles in about 6,650 journals in science and engineering¹¹. Co-authorship counts use the street address associated with the institutional affiliation of each author of an article.

¹⁰ Some recent developments in the world city network literature offer exploratory theoretical grounding using actor-based random utility modelling (see Liu et al., 2013).

¹¹ We recognize that the Science Citation Index is just one way of measuring scientific collaboration and output. It is also possible to use citations or patents, although the latter involves a different literature and a different set of challenges (see Hu, 2010). We believe that both citation and, especially, patent data are associated with serious problems as regards the identification of *actual* spatial locations of the relevant inputs. Nevertheless, we acknowledge that the use of the SCI index comes with the biases inherent in this type of data. For instance, Hennemann et al. (2011) points out that there can be differences when using domestic bibliographic databases as opposed to international ones. However, given that science policy-makers attach increasing value to the *international* visibility of research, we believe that our focus on the rise of Chinese science as measured by SCI publication counts is justified (Jonkers, 2010, p.13). An additional advantage is that an international database allows for direct comparisons between cities and networks in different parts of the world.

The spatial delimitation of each city approximates labor market areas, thus including both a central city and its outlying suburbs. Comparable studies combine neighboring cities if the center-to-center time distance is less than 45 minutes (Matthiessen et al, 2002). In the Chinese case, the labor market areas in practice correspond to the urban districts of each included city, since the administrative delimitations of Chinese cities encompass rural hinterlands. China's universities and research institutes tend to have central urban locations, although suburban "university districts" are becoming more common. Even so, there are no Chinese college towns as remote as College Station (Texas), Ithaca (New York) or State College (Pennsylvania).

The 31 included cities account for almost all SCI-indexed publications from mainland China (see Table 2). The covered time periods are 1996-1998, 2002-2004 and 2008-2010. These three-year periods are consistent with the network analyses of world science cities by Matthiessen et al. (2010), making international comparisons possible.

There was little Chinese scientific output and few co-authored papers in the first of the three periods. The 1996-to-1998 period is thus an "embryonic" period, followed by a period of middling scientific growth between 2002 and 2004. The final period—from 2008 to 2010—represents full-fledged growth involving much larger output quantities. These contrasts should enable us to capture the various spatial changes occurring within the Chinese science system.

4.2. Why use cities as the unit of analysis?

This study is part of the "scientific collaboration in space" literature (see Frenken et al., 2009). What makes our work distinct is its focus on scientific collaboration *between cities*. There are a few reasons for this. First, science production clusters in places such as large cities and university towns and is more spatially concentrated than most types of production (Liefner and

Hennemann, 2011; Matthiessen et al, 2002). Second, decreasing communication costs have not caused the obsolescence of cities. Instead, cities have reinforced their importance as coordination centers of spatially dispersed activities (Sassen, 1991). Third, cities do not exist in isolation: they are nodes in interconnected networks (Taylor, 2004). Fourth, functional urban regions correspond to integrated labor markets, including markets for scientific labor. Fifth, cities are better suited for the questions at hand—such as spatial political bias. For example, a “same-province effect” necessitates a separation between intra-provincial and inter-provincial links.

4.3. Gravity estimation

Scientific interaction is bidirectional. Unlike in trade, there is no obvious source or destination region when two or more scientists co-author a paper. Adaptations of the gravity model for the purpose of analyzing scientific cooperation should therefore include only one volume or mass variable, typically the product of the total publication volumes of two regions. Additionally, co-authorships consist of non-negative integer values (count data), which render ordinary-least-squares estimation inappropriate (Hilbe, 2011). Estimation of gravity models of co-authorships such as equation (4) thus tend to be based on a Poisson process:

$$\Pr[C_{ij}] = \frac{\exp(-\mu_{ij})\mu_{ij}^{C_{ij}}}{C_{ij}!}, \quad \mu_{ij} = \exp(\beta_1 + \beta_2 \ln M_i M_j + \beta_3 \ln D_{ij}) \quad (5)$$

where scientific collaboration (C_{ij}) between cities i and j follows a Poisson distribution with conditional mean μ . The mass variable ($M_i M_j$) and distance variable (D_{ij}) are dependent on this conditional mean.

An alternative count model is the negative binomial model. The key difference concerns the conditional variance. The Poisson regression model assumes a Poisson distribution, where the conditional mean of the dependent variable equals the conditional variance. It is however common for the conditional variance to exceed the conditional mean, especially when the count variable has more zeroes than a Poisson-distributed data-generating process would yield. The resulting over-dispersion can be accounted for in the negative binomial model via an extra parameter denoted as α (Hilbe, 2011). In the case of Chinese co-authorships, an over-dispersion test consistently rejected the null hypothesis that the conditional variance of the dependent variable equals its conditional mean. Thus, the gravity-type regressions employ the negative binomial regression model¹².

Table 3 describes and lists the sources of the dependent and independent variables in the gravity-type models. Our main mass variable (PUBMASS) follows the empirical literature and is the product of the SCI publication counts of two cities. As robustness checks we also consider alternative proxies for the mass variable: city publication products two years before the start of the observed co-authorship period (PUBMASS2); the product of the cities' GDP (GDPMASS); the product of the cities' volumes of tertiary teachers (TEACHMASS); and the product of the number of national universities in each of the two cities (UNIMASS). GDPMASS is often included in trade models and proxies for overall agglomeration economies. TEACHMASS and UNIMASS are proxies for investments in scientific research.

The motivation behind these variables is threefold. First, the alternative measures address potential endogeneity problems associated with PUBMASS. We adopt lagged values to address potential simultaneity problems, while random effects and time dummies control for omitted variables that may be important¹³. Second, TEACHMASS and UNIMASS are

¹² See Scherngell and Hu (2011) for a formal presentation of the negative binomial model that is applied to scientific flows.

¹³ Acosta et al. (2011) is the only study to address the first of these issues apart from the present study.

potential input variables in line with Beckmann's gravity model. Third, a comparison of the effects of UNIMASS and GDPMASS may offer clues regarding the relative importance of politics-driven and market-driven interaction processes. The former variable reflects investments in science, while the latter reflects overall agglomeration economies, including potential knowledge externalities from industry.

TABLE 3 ABOUT HERE

5. Results

5.1. Descriptive analysis

We present the number of co-authorships of each city with the other 30 cities in Figure 1. Beijing, Shanghai and Nanjing are top collaboration centers vis-à-vis other cities in all three periods. Guangzhou, Wuhan and Hangzhou are also important, with Guangzhou and Hangzhou replacing Hefei and Shenyang from 2002 onward. The least important centers of co-authorships are roughly the same across all periods and include Guiyang, Nanchang, Nanning, Ningbo, Shijiazhuang and Xiamen. Shenzhen is exceptional in that it became a much more interactive node between the second and third periods.

FIGURE 1 ABOUT HERE

One way of looking at China's network structure is to identify the most intensive co-authorship links. Figures 2 and 3 show all links involving .25 percent or more of the total paper output of the 31 cities in 1996-1998 and 2008-2010, respectively. In the earlier period, this implies at

least 150 co-authored papers, while in the later period the cut-off point is 1,000 papers. The increase reflects the rapid growth in China's science output.

What emerges is a Chinese science network that is Beijing-centric. In the earlier period, there were a total of twelve high-frequency links, eleven of which involve Beijing. In the later period, the total number of links had increased to 21 links, with 18 Beijing links. Since the identification of links is in relation to the total production of scientific papers in the relevant time period, the results show that intercity co-authorships increased in relative as well as absolute importance. China's science network became more interactive.

Figures 2 and 3 also hint at different hierarchical levels in the Chinese science network. The relevant criterion is whether a city has one or more high-frequency links to other cities. In the earlier period, Beijing had such links with 11 cities. These 11 cities were all among the top 17 in total science output, and all are provincial capitals. In the second period, the second level encompassed 18 of the 19 largest Chinese science cities after Beijing in SCI output terms. The second period also exhibited what may prove to be two emergent tendencies: two non-Beijing links in the Yangtze River Delta region as well as two links between Beijing and cities that are not provincial capitals (Dalian and Qingdao).

FIGURE 2 ABOUT HERE

FIGURE 3 ABOUT HERE

5.2. Gravity results

Table 4 gives the results of estimated pooled negative binomial regressions with five different specifications. The five regressions cover all three time periods and include the logarithm of

the product of the SCI publication volumes in cities i and j as well as the logarithm of the distance in kilometers between i and j .

TABLE 4 ABOUT HERE

Model (1) is the basic model with only mass and distance, while models (2) through (5) introduce various refinements. Models (2) and (3) introduce city fixed effects, while model (4) employs Beijing and same-province dummies instead of 30 fixed effects. Models (3) and (4) use time dummies to account for time trends. Model (5), finally, uses combined *city-and-time* fixed effects.

As expected, an increase in the product of the total number of publications is associated with an increase in the number of co-authored papers. The coefficient estimates range from .63 to .84, which is in line with comparable prior studies (Ponds et al, 2007; Scherngell and Hu, 2011). The distance effects have the expected negative sign and are highly significant. The estimated magnitudes are all in the vicinity of -.30. Most comparable studies report distance effects between -.23 and -.70, implying that spatial friction is not a greater impediment to interaction in China than in the West (see Hoekman et al, 2009; Scherngell and Hu, 2011). We may interpret this result as indirect evidence that China's transport and communication infrastructures are unusually advanced for a middle-income country.

Model (4) shows that there are more interactions with Beijing than the output volume of the city *and* its geographic location vis-à-vis other Chinese cities jointly imply. This result corroborates Beijing's role as a top-level coordination center in the network. That a Beijing link should be attractive to scientists in other localities is unsurprising; the Chinese Academy of Sciences is in Beijing as are China's two global top-50 universities (Peking University and

Tsinghua)¹⁴. Consequently, Beijing-based scientists receive a disproportionate share of science funding (Feng and Pei, 2011). This is the first result in support of the first hypothesis regarding spatial political bias, which is that Beijing is over-represented in co-authorships even after having accounted for its over-representation in China's science output, given the size of Beijing's economy relative to other Chinese cities.

There is also a significant same-province effect, implying more intra-provincial collaboration than volume and spatial proximity considerations would lead us to expect. This supports the hypothesis of a same-province bias, even after controlling for the spatial proximity of same-province cities.

Likelihood ratio tests indicate that models (1), (2) and (4) are nested in models (3) and (5), whereas (3) is not nested in (5). In other words, models with city fixed effects and time dummies—entered either separately or jointly—outperform simpler regressions. These models are similar to Hoekman et al.'s (2010) models of inter-regional scientific interaction in Europe. Their estimated distance coefficient for aggregate science equaled $-.57$ for 2000 to 2007, after controlling for regional, national and linguistic border effects. There is thus some evidence that spatial friction may have a *greater* inhibitory effect in Europe than in China. This might suggest that national boundaries affect cross-border scientific cooperation in ways that are not easy to control for in formal models (Okubo and Zitt, 2004).

TABLE 5 ABOUT HERE

TABLE 6 ABOUT HERE

¹⁴ Such links may partly be explained by the fact that some institutes of the Chinese Academy of Sciences often fund research and have links with institutes in other cities.

Table 5 decomposes the analysis into the three time periods, using cross-sectional models. The estimated variable coefficients show a great deal of consistency from one period to the next, with significant mass and distance effects. The distance effect first increases and then decreases over this time period. One possible explanation is that there are two opposing effects at work: one where scientists in peripheral cities increasingly enter into long-distance research relationships; and one that reflects decreasing spatial friction as a result of investments in interregional transport infrastructure. A decomposition of these potential opposing tendencies is however infeasible with the available data.

The functions with Beijing and same-province dummies show that the Beijing effect increases over time. Though other cities are increasing their contributions to China's total science output, there is also an increasing tendency—in relative terms—for co-authored papers to involve Beijing-based scientists. There are thus indications that the top-level monocentricity of the spatial political bias increased between 1996 and 2010. The increasing “network monocentricity” is not necessarily intentional; it might be an unintended consequence of the tendency of scientists from smaller cities to seek out collaboration partners in the core so as to gain greater resource access (Acosta et al., 2011).

The same-province dummy estimates imply that the provincial aspect of the spatial political bias is also increasing. China thus offers a stark contrast to the evolution of the European science network as estimated by Hoekman et al. (2010). While national borders are becoming less important impediments to European scientific cooperation, our results imply that provincial borders are becoming more important in China. Hence, there are indications that regional protectionism not only exists at the provincial level, but also that it may be increasing over time. Such protectionism may constitute the main impediment to future integration (Yoon, 2011), since falling space-bridging costs imply that spatial friction is becoming less of an obstacle to inter-provincial scientific interaction.

The final set of regressions consists of negative binomial panels with random effects. The panels introduce four new specifications with exogenous mass variables along with the original one. All four regressions serve as robustness checks, with the last two being closer to Beckmann's input model (TEACHMASS and UNIMASS) and a GDPMASS vs UNIMASS comparison relating to the markets-versus-politics dichotomy.

Table 6 presents the results. The results suggest that there are many ways of approximating mass, and that the distance effect is relatively robust across specifications. The more exogenous measures of mass are all robust predictors of scientific interaction, while a critical look at the popular but potentially more endogenous measure (PUBMASS) suggests that it may overestimate the size effect. A comparison between GDPMASS and UNIMASS reveals that UNIMASS is the more significant predictor of scientific collaboration. This again hints at the importance of political resource allocation in science as opposed to market-driven agglomeration effects. Outside China, such agglomeration effects often dominate political allocation, as is attested by the dominant role of New York in financial research and the San Francisco Bay Area in IT-related science and engineering. San Francisco, not Washington, is America's largest and most connected science city.

6. Understanding the gravity results: an exploratory analysis

A simple descriptive network analysis is helpful for identifying which particular inter-city linkages are most important for the Beijing and same-province effects. One way of revealing strong links is to identify whether links are stronger or weaker than expected in a relative sense, controlling for mass, *but not for distance*. Thus we should expect all stronger-than-expected links to consist exclusively of the most geographically proximate cities, in the absence of spatial political biases.

If there were no spatial friction, the expected number of co-authorships involving city i and city j should equal the product of the number of co-authored articles (with all cities in the network) in the two cities divided by the total number of network co-authorships. Figure 4 shows those links where the observed number of co-authorships exceeds 150 percent of the expected number. Lines of medium thickness imply between two and three times as many observed as expected co-authorships, while the thickest lines imply observed frequencies that are over three times greater than expected. Table 7 shows the ten links with the highest ratios between observed and expected co-authorships. It is the identification of disproportionately strong links that makes it possible to construct proximity clusters such as Taylorite “bands.”

Figure 4 reveals two tendencies. First, Beijing’s strongest links are with cities with *small* SCI publication volumes, including geographically remote ones such as Kunming and Guiyang. Its six disproportionately strong links are with cities that are ranked 10th, 17th, 20th, 25th, 28th, and 31st in output (see Table 2). Another aspect of this tendency is that the smallest science cities tend to have numerous strong links. Guiyang, Kunming, Nanchang, Nanning, Ningbo, Shijiazhuang, Suzhou, and Taiyuan have three or four strong links each. It is likely that this reflects intercity cooperation strategies among universities in cities with small local knowledge stocks. This strategy should be especially attractive in China, since scientists working at top-ranked national universities have better access to national research funds.

FIGURE 4 ABOUT HERE

TABLE 7 ABOUT HERE

The second tendency is proximity clustering in regional bands, with the strongest links being intra-provincial. There are six intra-provincial pairings of cities, four of which are among the

ten links with the greatest observed-to-expected ratios (see Table 7). The remaining two intra-provincial links are also stronger than expected.

Each of the eight city pairs in bold font in Table 7 represents a link between cities in the same part of China, which implies proximity clustering. Although short geographic distances may be a partial explanation, both Figure 4 and Table 7 hint at an independent provincial effect. In the Yangtze River Delta region, Suzhou in Jiangsu is more strongly linked to Nanjing (Jiangsu) than to Shanghai, though the latter city is nearer. Hangzhou in Zhejiang links to Ningbo (Zhejiang) but not to equidistant Suzhou. And Nanjing does not have a strong link to Hefei. This could be an indication that second-ranked provincial cities seek linkages to the provincial capital, which host national universities as well as provincial funding organizations. A conjecture is therefore that China's science network consists of at least three levels. There is also the possibility that each province makes up a band of closely connected cities, with hierarchical control functions vested in the provincial capital. A likely consequence of this province-level home bias is that it impedes effective knowledge transfer across China (Chen and Wang, 2003; Hui, 2007; Yoon, 2011).

The Guangzhou-Jinan and Shenzhen-Harbin links may seem puzzling. These links reflect a new expansion strategy among some cities with modest science agglomerations. Several such cities have pursued scientific cooperation with other cities so as to attain rapid scientific growth. Important means include national initiatives such as the 973 and 211 programs. The 973 program funds large-scale research projects, including topic-specific projects involving universities in several Chinese cities. The 211 program is even more important, since it receives most Chinese research funding and includes most national key laboratories. These laboratories sometimes encompass universities in different provinces (Hu

and Jefferson, 2008; Jonkers, 2010, p.25)¹⁵. The UNIMASS variable in Table 6 is a measure of the impact of the 211 program, since it measures the number of national universities funded by it.

Shenzhen's expansion involves more than these programs, however. In 2002, the city government decided to build a university zone and encouraged national universities from elsewhere to set up graduate schools in Shenzhen. The Harbin Institute of Technology is the most prominent such school. This could explain how Shenzhen has not only progressed rapidly in scientific collaboration after the 2002-2004 period but also why the Shenzhen-Harbin link is so strong. Though unusual, the strong Guangzhou-Jinan and Shenzhen-Harbin linkages both owe their strength to political rather than market interactions.

Overall, the network patterns suggest a strong connection between the spatial structure of Chinese science and political resource allocation. Beijing dominates nationally, and provincial capitals dominate provinces. All top universities are public, and research funds are governmental. Rich cities with a strong market orientation but less political power—such as Ningbo, Shenzhen and Suzhou—have modest publication volumes, albeit with high growth rates. The smaller science cities are growing, but the deck is still stacked in Beijing's favor.

FIGURE 5 ABOUT HERE

Figure 5 shows an exploratory Chinese system of science cities¹⁶. It depicts China's publication and co-authorship structure along the lines of earlier studies of world science cities; cities thus

¹⁵ We thank a referee for pointing this out to us. Our own search of the Ministry of Education and Ministry of Science and Technology sources revealed that Sun Yat-Sen University (Guangzhou-based) and Shandong University (Jinan-based) have well-established links and jointly participate in a major research project that also includes Nankai University, Peking University and Hangzhou Normal University.

¹⁶ Beijing has much greater volume and network connectivity than other cities and is the only top-level city. We used quantitative thresholds to allocate cities to levels two through four. The measured quantities refer to the most recent period (2008-2010). During that period, level 2 cities generated at least 15,000 SCI papers, of which at least 6,500 were co-authored together with authors from the other 30 Chinese cities. Level 3 cities have at least 8,000 SCI papers *and* 4,000 intercity co-authorships. There is also a third criterion, which reflects the

belong to hierarchical levels and possibly also to bands of strong links. The uppermost level consists of Beijing alone, with the second-tier cities being Shanghai, Nanjing, Wuhan, Guangzhou and Hangzhou. A higher hierarchical level implies greater overall network connectivity. On the whole, the picture that emerges is a monocentric network.

There are several instances of regional bands involving cities at different levels. The indicated bands are the North (led by Beijing), the Yangtze River Delta (Shanghai), the Northeast (non-hierarchical), the Center (Wuhan); and the West (Chengdu).

7. Conclusion

Chinese science is on the rise. This paper explores the geography of Chinese science, with a focus on the role of China's spatial political bias, which manifests itself as Beijing and same-province effects. Gravity-type models reveal that mass and distance variables conform to the global pattern, but with slightly less spatial friction and with an orientation towards co-authorships involving Beijing-based scientists. The relatively low level of spatial friction alludes to large center-periphery distances and well-functioning transport and communication infrastructures for intercity flows of people and ideas.

The presence of a same-province effect implies lower-level spatial political bias. Moreover, both the Beijing and the home-province biases became more pronounced between 1996 and 2010, which is the opposite of what happened in Europe.

Our exploratory network analysis found a spatial organization of Chinese science where Beijing alone occupies the highest hierarchical level. Though there is scientific growth elsewhere in the system, Beijing continues to dominate. This analysis also hints at the existence

diversification of a city's intercity connectivity. The third criterion conforms to the "world network of science cities" literature, where cities closer to the innermost concentric circle have more "global" connectivity. The criteria explain why both Xi'an and Jinan are at level 3 rather than level 2: Xi'an produces more than 15000 papers but is less connected and spatially diversified than Jinan, while Jinan's overall publication output is too small (see Table 2 and Figure 1).

of “regional bands” of cities. Unusually strong intra-provincial links reflect the regional dimension of China’s spatial political bias.

Overall, our results show that political decisions rather than markets shape the spatial resource allocation of scientific resources. In a market-led spatial organization of science, one would expect a city’s scientific output and connectivity to reflect general knowledge-related agglomeration economies. Not so in China, where Beijing is twice as large as Shanghai and 30 times larger than Shenzhen, and with even greater differentials in terms of network connectivity. We think that this bias reflects the absence of a true market for ideas in China. Coase and Wang (2013) contend that China’s central planning of ideas (including science) is its greatest economic problem. They write that “Chinese universities have become more skilled in currying favor with the Ministry of Education than in offering innovative research and educational progress; this situation is not so different from state-owned enterprises before reform. (...) As far as Chinese universities are concerned, the government essentially controls both inputs (finance and personnel) and outputs (degree programs), leaving them very little autonomy” (pp. 191-92).

While China has arrived on the world stage of science, it still occupies a peripheral position in the production of new scientific breakthroughs. Even Beijing lacks the impact of Western cities with similar publication volumes, such as London or Paris (Nature, 2013; Royal Society, 2011). The main challenge will be to break out of this isolation. The problem is all the more daunting since it may be more cultural than political. It is a problem that mainland China shares with Japan, South Korea and Taiwan (Andersson et al., 2013). When scientists around the world look for path-breaking ideas they still prefer Cambridge—whether in England or Massachusetts—over Beijing or Tokyo. To break out of this impasse, China may need a new “cultural revolution” more than it needs money.

References

- Acosta M, Coronado D, Ferrándiz E, León M D, 2011, “Factors affecting inter-regional academic scientific collaboration within Europe: the role of economic distance” *Scientometrics* **87** 63-74
- Andersson Å E, Andersson D E, 2006 *The Economics of Experiences, the Arts and Entertainment* (Edward Elgar, Cheltenham)
- Andersson Å E, Persson O, 1993, “Networking scientists” *Annals of Regional Science* **27** 11-21
- Andersson Å E, Andersson D E, Matthiessen C W, 2013 *Öresundsregionen: den dynamiska metropolen* (Dialogos, Stockholm)
- Beckmann M, 1993, “Knowledge networks: The case of scientific interaction at a distance” *Annals of Regional Science* **27** 5-9
- Beckmann M, 1994, “On knowledge networks in science: collaboration among equals” *Annals of Regional Science* **28** 233-42
- Beckmann M, 1999 *Lectures on Location Theory* (Springer, Berlin)
- Chen C T, Wang T, 2003, “The influence to the local S&T legislation from the view of CAST operation mode forum on science and technology in China”, *Forum on Science and Technology in China* **4** 3–6 [in Chinese]
- Coase R, Wang N, 2013 *How China Became Capitalist* (Palgrave Macmillan, Basingstoke)
- Feng Z H, Pei T C, 2011, “The ‘10 billion’ cake: National Natural Science Foundation of China” *Journal of Thoracic Disease* **3** 213-216
- Frenken K, Hardeman S, Hoekman J, 2009, “Spatial scientometrics: towards a cumulative research program” *Journal of Informetrics* **3** 222-232
- Hansen T, 2013, “Bridging regional innovation: cross-border collaboration in the Øresund Region” *Geografisk Tidsskrift-Danish Journal of Geography* **113**(1) 25-38

- Havemann F, Heinz M, Kretschmer H, 2006, "Collaboration and distances between German immunological institutes – a trend analysis" *Journal of Biomedical Discovery and Collaboration* **1** 1-7
- Hennemann S, Wang T, Liefner I, 2011, "Measuring regional science networks in China: a comparison of international and domestic bibliographic data sources" *Scientometrics* **88**(2) 839-857
- Hilbe J M, 2011, *Negative Binomial Regression* 2nd edition (Cambridge University Press, Cambridge)
- Hoekman J, Frenken K, Tijssen R J W, 2010, "Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe" *Research Policy* **39** 662-673
- Hoekman J, Frenken K, van Oort F, 2009, "The geography of collaborative knowledge production in Europe" *Annals of Regional Science* **43** 721-738
- Hu A G Z, 2010, "Propensity to patent, competition and China's foreign patenting surge" *Research Policy* **39** 985-993
- Hu A G Z, Jefferson, G H, 2008, "Science and Technology in China" in *China's Great Economic Transformation* Eds L Brandt, TG Rawski (Cambridge University Press, Cambridge) pp 286-336
- Hui L, 2007, "Sub-national Innovation System Practices in China" Paper presented at the national workshop on sub-national innovation systems and technology capacity building policies to enhance competitiveness of SMEs, organized by UN-ESCAP and Indonesian Institute of Sciences (LIPI), 3 – 4 April 2007, Jakarta, Indonesia
- Isard W, 1960, *Methods of regional analysis: an introduction to regional science* (MIT Press, Cambridge, MA)
- Jonkers K, 2010 *Mobility, Migration and the Chinese Scientific Research System* (Routledge, London)

- Katz J S, 1994, “Geographical proximity and scientific collaboration” *Scientometrics* **31**(1) 31-43
- Liang L, Zhu L, 2002, “Major factors affecting China’s inter-regional research collaboration: regional scientific productivity and geographical proximity” *Scientometrics* **55** 287-316
- Liefner I, Hennemann S, 2011, “Structural holes and new dimensions of distance: the spatial configuration of the scientific knowledge network of China’s optical technology sector” *Environment and Planning A* **43** 810-829
- Liu X, Derudder B, Liu Y, Witlox F, Shen W, 2013, “A stochastic actor-based modelling of the evolution of an intercity corporate network” *Environment and Planning A* **45**(4) 947 – 966
- Marginson S, 2011. “Higher Education in East Asia and Singapore: Rise of the Confucian Model” *Higher Education* **61**, 587 – 611
- Matthiessen C W, Schwarz A W, Find S, 2002, “The top-level global research system, 1997-99: centres, networks and nodality. An analysis based on bibliometric indicators” *Urban Studies* **39** 903-927
- Matthiessen C W, Schwarz A W, Find S, 2010, “World cities of scientific knowledge: systems, networks and potential dynamics. An analysis based on bibliometric indicators” *Urban Studies* **47**(9) 1879-1897
- Matthiessen C W, Schwarz A W, Find S, 2011, “Research nodes and networks,” in *Handbook of Creative Cities* Eds D E Andersson, Å E Andersson, C Mellander (Edward Elgar, Cheltenham) pp 211-228
- Nature, 2013 *Nature Publishing Index 2012: China* (Nature Publishing Group).
- Okubo Y, Zitt M, 2004, “Search for research integration across Europe: a closer look at international and inter-regional collaboration in France” *Science and Public Policy* **31**(3) 213-226

- Oner A C, Mitsova A, Prosperi D, Vos J, 2010, "Knowledge globalization in urban studies and planning: a network analysis of international co-authorships" *Journal of Knowledge Globalisation* **3** 2-30
- Ponds R, van Oort F, Frenken K, 2007, "The geographical and institutional proximity of research collaboration" *Papers in Regional Science* **86** 423-443
- Royal Society, 2011 *Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century* (Elsevier, Amsterdam)
- Sassen S, 1991 *The Global City: New York, London, Tokyo* (Princeton University Press, Princeton, NJ)
- Scherngell T, Hu Y, 2011, "Collaborative knowledge production in China: regional evidence from a gravity model approach" *Regional Studies* **45** 755-772
- Sen A, Smith T E, 1995, *Gravity Models of Spatial Interaction Behavior* (Springer, New York)
- Taylor P J 2004 *World City Network: A Global Analysis* (Routledge, London)
- White H D, 2011 "Scientific and Scholarly Networks" in *The SAGE handbook of social network analysis* Eds J P Scott, P Carrington (London: Sage Publications) 271-285
- Wilson A G, 1970 *Entropy in Urban and Regional Modelling* (Pion, London)
- Yoon J, 2011, *Exploring regional innovation capacities of PR China: Toward the study of knowledge divide* (PhD thesis, Georgia Institute of Technology)
- Zhou P, Leydesdorff L, 2006, "The emergence of China as a leading nation in science" *Research Policy* **35**(1) 83-104
- Zhou P, Thijs B, Glänzel W 2009, "Regional analysis on Chinese scientific output" *Scientometrics* **81**(3) 839-857