

Les villes et la science : apports de la spatialisation des données bibliométriques mondiales

Denis Eckert, Myriam Baron, Laurent Jégou

Résumé.— Les données bibliométriques sont l'une des principales sources pour l'étude de la production scientifique mondiale. Or les géographes ont jusqu'ici peu participé au travail sur ces données. Celles-ci permettent pourtant de caractériser l'activité scientifique des villes, à une échelle d'analyse particulièrement pertinente. En s'appuyant sur le *Web of Science*, on identifie les dimensions géographiques des informations bibliométriques. On propose une méthode de géocodage, puis d'agrégation spatiale des données primaires qui donne sens à la notion d'« agglomération scientifique » et autorise de nombreux traitements.

Bases de données • Géovisualisation • Matrices d'information spatiale • Publications scientifiques • Villes

Abstract.— **Cities and science: the contributions of the spatialisation of global bibliometric data.**— Bibliometric data is one of the main sources for the study of global scientific production and yet, geographers have yet to participate fully in work on such data. Such information allows the scientific activity of cities to be characterized at a particularly relevant scale of analysis. Using the Web of Science, we identify the geographic dimensions of bibliometric information. We then outline a method of geocoding and then of spatial aggregation of primary data, which gives meaning to the notion of "scientific agglomeration" and allows many treatments.

Cities • Databases • Geovisualization • Matrices of Spatial Information • Scientific Publications

Resumen.— **Las ciudades y la ciencia: aportes de la espacialización de los datos bibliométricos mundiales.**— Los datos bibliométricos conforman una de las principales fuentes para el estudio de la producción científica mundial. Ahora bien, los geógrafos participaron poco a su explotación. Sin embargo, estos datos permiten caracterizar la actividad científica de las ciudades con una escala de análisis particularmente pertinente. Apoyandose en el Web of Sciences, se identifican las dimensiones geográficas de las informaciones bibliométricas. Se propone un método de geocodificación, luego de agregación espacial de los datos primarios, el cual da sentido a la noción de "aglomeración científica" y autoriza a numerosos tratamientos.

Banco de datos • Ciudades • Geovisualización • Matrices de información espacial • Publicaciones científicas

Introduction : recherche, productions scientifiques et enjeux spatiaux

Le géographe, qui souhaite appréhender l'activité de recherche et concevoir des indicateurs spatialisés pertinents pour en rendre compte, doit prendre en considération le fait que les données relatives au nombre de chercheurs, à leurs activités, mais aussi à leurs coopérations et aux financements qu'ils obtiennent, se révèlent difficiles à obtenir, encore plus à comparer. La diversité des institutions qui

abritent une activité de recherche rend ce dénombrement très aléatoire d'un pays à l'autre. En fonction des définitions et de la qualité des comptages nationaux, on trouve ici des cohortes et là quelques individus isolés. Mener des comparaisons internationales à partir de telles informations est bien hasardeux.

Mesurer l'activité des chercheurs constitue un autre défi. L'analyse de l'activité de recherche est une branche de la géographie de l'innovation (Feldman *et al.*, 2002). D'importants travaux ont été menés sur des instruments de financement et de coopération aux niveaux national ou international (Comin, 2009), ou sur les brevets déposés par les laboratoires ou les individus (Massard *et al.*, 2003; Berroir *et al.*, 2009), qui peuvent être considérés comme autant de mesures de l'innovation produite par les chercheurs. Ces travaux ont néanmoins des limites liées à la nature même de leurs sources: le périmètre institutionnel des instruments de financement et de coopération. On ne peut alors que mesurer l'activité des chercheurs telle qu'elle se déploie dans ces cadres construits *a priori*, et qui ne couvrent pas l'ensemble de leur domaine d'intervention. Une étude de la recherche financée par l'Union européenne restera, par construction, limitée à ce périmètre précis.

C'est pourquoi, dans le champ des *Science Studies*, on s'est beaucoup intéressé au produit le plus visible de l'activité des chercheurs: les publications. Des sociologues de la science, mobilisant les ressources de la bibliométrie, se sont appuyés sur les informations contenues dans les bases bibliographiques internationales pour mesurer et analyser les productions scientifiques à l'échelle mondiale. Ils ont ainsi procédé à un véritable détournement du contenu de ces bases, en adoptant une approche bibliométrique à des fins analytiques — ce que les concepteurs n'avaient pas prévu (Wouters, Duval, 2006). Ils ont fondé une communauté thématique — la scientométrie — ayant développé sa propre méthodologie et créé ses propres revues (*Scientometrics* ou *JASIST, Journal of the American Society for Information Science and Technology*). Ces travaux s'appuient sur l'hypothèse que la publication est l'une des activités caractéristiques des chercheurs, si ce n'est la principale finalité de leur travail, comme Bruno Latour et Steve Woolgar le constataient dès 1979: «les acteurs reconnaissent que la production d'articles est le but essentiel de leur activité». L'examen des articles référencés dans ces bases de données est ainsi susceptible de nous en apprendre beaucoup sur le fonctionnement social de la science. Et ceci même si les biais de ces bases sont évidents. Ils ont d'ailleurs été maintes fois discutés (1). On peut remarquer que les géographes ont, dans l'ensemble, peu investi ce champ d'étude, resté pour l'essentiel l'apanage des sociologues de la science ou des économistes de l'innovation. Or il semble possible de développer substantiellement l'analyse géographique globale et multiniveau du fonctionnement de la science par l'étude des bases de données bibliographiques. Le pari est d'arriver à structurer l'information de manière à pouvoir analyser spatialement l'activité scientifique.

L'objectif de cet article est de préciser en quoi la nature des données de publications permet un codage géographique fin, ce qui autorise une multiplicité de traitements, articulant tant l'ensemble des échelles spatiales que la dimension temporelle. Il faudra s'intéresser aux modalités d'accès à ces données, collectées et maintenues par des grands opérateurs internationaux (Thomson-Reuters, Elsevier).

Nous montrerons ensuite que des chercheurs se sont déjà saisis des problématiques spatiales, et que les acquis sont significatifs en ce domaine, comme

l'explique bien Koen Frenken dans son article sur les « Spatial Scientometrics » (2009). En revanche, la prise en compte d'un niveau fondamental pour l'analyse territoriale de la science — le niveau urbain — reste incomplètement réalisée. L'essentiel des travaux s'appuie sur la petite échelle (internationale) ou les échelles intermédiaires (région, département ou équivalent), tandis que les recherches portant sur les villes elles-mêmes sont rares. Or les productions scientifiques ont tout à gagner à être analysées dans leurs dimensions urbaines. Comme il a été maintes fois rappelé au cours des 50 dernières années, l'université, qui regroupe une grande partie des activités de formation supérieure et de recherche, est et demeure un équipement urbain ; même si cette nature urbaine peut renvoyer à des rapports à la ville variés, traduits en partie par des implantations aussi différentes que le quartier universitaire en centre-ville et le campus en périphéries plus ou moins éloignées (Vassal, 1969 ; Frémont, 1990 ; Brunet, 1990 ; Frémont *et al.*, 1992). Prendre en compte cette, voire, ces dimensions urbaines, c'est aussi s'autoriser à analyser l'inscription multiniveau de la science dans les systèmes urbains, afin de pouvoir caractériser le rôle des villes dans ce domaine, aux niveaux national, régional et bien entendu international.

Il y a donc un enjeu autour du codage à un niveau géographique fin des données de publication scientifique. Mais ce codage pose de multiples problèmes, ce qui explique entre autres les difficultés rencontrées par maints spécialistes pour traiter les questions spatiales à une échelle fine. Nous proposerons donc une technique de géocodage puis d'agrégation des données qui permet de construire des entités spatiales cohérentes pour l'ensemble de la base Thomson-Reuters, dénommée *Web of Science (WoS)*, et ainsi de prendre en compte la dimension urbaine dans l'analyse de l'ensemble du système mondial de la science.

1. Dimension géographique des données bibliographiques

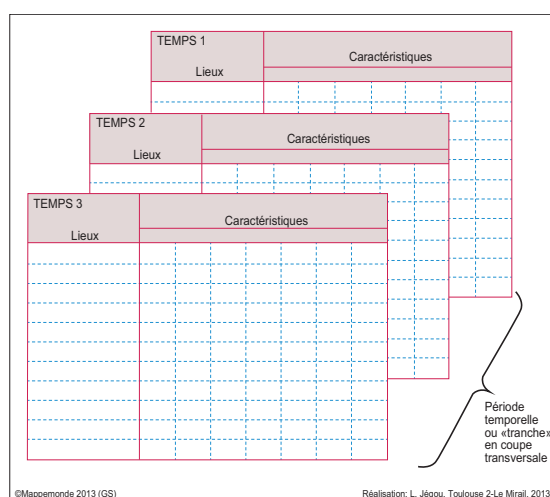
Le *Web of Science* est la plus ancienne des bases de données mondiales sur la science. Créée dans les années 1960 (elle était alors constituée du seul *Science Citation Index*), elle s'est étoffée au fil du temps, couvrant toujours plus de domaines scientifiques et de revues. Le *WoS* s'affranchit progressivement de son « péché originel » d'avoir recensé, dans les premiers moments de son existence, essentiellement la production scientifique nord-américaine. Il a, en effet, élargi plusieurs fois son « périmètre », intégrant de mieux en mieux les revues européennes à partir des années 1990. Actuellement, le *WoS* indexe presque 12 000 revues. Dans le même temps, les chercheurs du monde entier ont eu tendance à internationaliser leurs pratiques de publication, ciblant davantage les nombreuses revues indexées. On peut donc considérer que les contenus du *WoS* sont de moins en moins biaisés du point de vue géographique, surtout depuis la fin des années 1980. *Scopus* est une base plus récente, lancée seulement en 2004. Dès le départ, ses concepteurs vantent sa très large couverture thématique et géographique. Aujourd'hui son volume est comparable à celui du *WoS*, même s'il ne remonte pas aussi loin dans le temps : 45 millions de notices en 2011 et quelque 15 000 revues indexées. Les progrès de l'indexation de l'activité scientifique par les opérateurs de ces bases de données donnent ainsi de plus en plus de légitimité aux approches géographiques qui s'appuient sur ces informations.

Les bases bibliographiques internationales présentent, en effet, un avantage considérable : on a affaire à une donnée normalisée internationalement. L'article de chimie préparé à Lyon est codé et repéré de la même façon que celui conçu à Mexico, s'il est publié dans une revue répertoriée par le *WoS* ou *Scopus* (2). On a donc un ensemble de notices qui constituent autant d'informations cohérentes sur les contenus d'un article : par le titre, la revue, les mots clés, le texte intégral, son (ses) auteur (s), son (leur) institution et son (leur) lieu d'exercice. Ce dernier élément permet de caractériser spatialement les contenus, et rend ainsi possible l'analyse de certains rapports de la production scientifique aux territoires (**encadré**).

Attributs spatiaux des notices du WoS							
<p>L'élément de base du <i>WoS</i> (ou de <i>Scopus</i>), autrement dit la brique à partir de laquelle il faut raisonner, est la notice. Cette donnée primaire recense pour un article donné :</p> <ul style="list-style-type: none"> • un (ou plusieurs) auteur(s); • leur(s) affiliation(s) institutionnelle(s); • leur(s) adresse(s); • la date de parution (année bibliographique); • le résumé, le titre de l'article, le nom de la revue, etc. <p>À partir de ces éléments, les informations localisées dans la base sont :</p> <ul style="list-style-type: none"> • un ensemble de lieux repérés par les adresses fournies par les auteurs; • le nombre de publications des équipes de ce lieu pour chaque année; • le nombre de relations entre lieux pour chaque année, révélé par les cosignatures d'articles; • les citations d'autres articles et ouvrages, puisque les références citées font aussi partie de la base et peuvent donc être localisées. <p>On a donc des données quantitatives. Elles deviennent spatialement sommables si l'on fait l'effort d'un comptage proportionnel* et que l'on « fractionne » une publication pour l'attribuer à un équipe de recherche et un lieu**. Par exemple, un article cosigné par 4 équipes (1 de Paris, 2 de Madrid, et 1 de Toulouse) ne peut compter qu'une fois dans la base globale : il est donc nécessaire de l'attribuer pour 1/4 à Paris, 1/2 à Madrid et 1/4 à Toulouse. La valeur « 1 » dans le <i>Web of Science</i> sera décomposée en :</p> <ul style="list-style-type: none"> • Toulouse et Paris : 0,25 chacun; • Madrid : 0,5. 		<p>Si l'on regroupe les données à l'échelle nationale, la France et l'Espagne compteront chacune pour 0,5. Ce procédé de fractionnement, assez simple dans son principe, permet donc toute agrégation ultérieure des données primaires*.</p> <p>L'unité spatiale élémentaire qui correspond à chaque ligne du tableau de base (fig. 1) correspond à la localisation la plus précise possible de l'adresse de l'auteur.</p> <p>On peut imaginer que les données relatives à un lieu soient exhaustives pour une année (toutes les publications attribuables à ce lieu, toutes les relations de ce lieu à d'autres lieux).</p> <p>Mais on peut évidemment procéder à toutes sortes de sélections et d'extractions en fonction de critères variables. On peut choisir de s'intéresser seulement à certains articles scientifiques, comme ceux issus de certaines revues ou correspondant à des critères de contenu particuliers, de s'intéresser, ou non, aux citations, aux cosignatures...</p> <p>*Il est particulièrement pertinent pour les variables « nombre de publications » et « nombre de cosignatures » ; car la somme globale de ces variables a un sens (nombre d'articles à l'échelle mondiale, nombre total de cosignatures). Il est moins important pour les citations.</p> <p>**Les bases ne permettent pas d'aller systématiquement jusqu'à l'auteur individuel et de le relier à un lieu : on raisonne donc en affiliations. L'entité primaire, « l'auteur » si l'on veut, est pour nous l'équipe de recherche avec son affiliation institutionnelle et son adresse, et non l'individu.</p>					
Unités spatiales les plus fines	Données ponctuelles	Données relationnelles (mesures d'intensité de relations de lieu à lieu)					
		Symétriques et non-orientées (si le lieu 1 a une relation avec le lieu α, la réciproque est vraie)			Orientées et asymétriques le fait qu'un article d'un auteur du lieu 1 cite un article d'un auteur du lieu α n'implique nullement la réciproque		
	Production	Cosignatures			Citations		
		P nombre (fractionné) de publications	Cosig _α nombre de cosignatures avec le lieu α	Cosig _β nombre de cosignatures avec le lieu β	...	Cit _α nombre l de citations d'un article fait dans le lieu α	Cit _β nombre m de citations d'un article fait dans le lieu β
lieu 1	P1	Cosig _α 1	Cosig _β 1	...	Cit _α 1	Cit _β 1	...
lieu 2	P2	Cosig _α 2	Cosig _β 2	...	Cit _α 2	Cit _β 2	...
lieu 3	P3	Cosig _α 3	Cosig _β 3	...	Cit _α 3	Cit _β 3	...

1.1. Du tableau élémentaire à la matrice d'information géographique

Les données bibliographiques du *WoS* et de *Scopus* sont donc riches en attributs spatiaux et susceptibles de contribuer à la création de matrices d'information géographique (Berry, 1964) au sens fort du terme (fig. 1). En effet, comme dans toute matrice d'information géographique, chaque cellule de ce premier tableau élémentaire, qui renvoie à la production, aux cosignatures ou encore aux citations, peut être additionnée à des cellules équivalentes: les trois variables de base (contribution, cosignature, citation) sont toujours sommables. Cela permet de procéder à des agrégations spatiales ou temporelles pour une ou éventuellement la totalité des variables, point qui se révèle d'une grande importance.



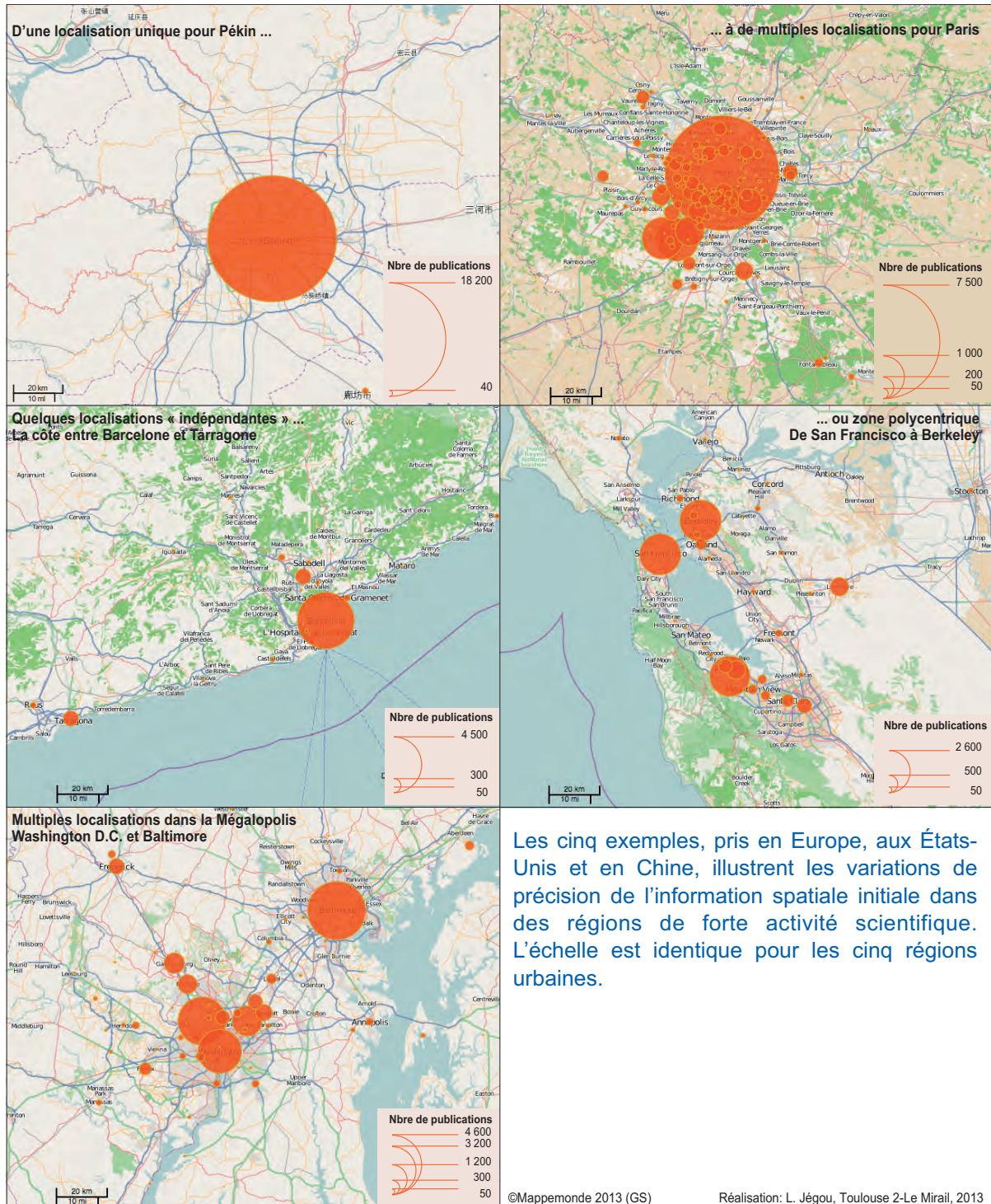
1. La matrice d'information géographique

Si l'on considère comme résolue la question de la localisation des adresses contenues dans la base, on peut imaginer que l'on a affaire à une matrice d'information géographique directement utilisable. Ce n'est malheureusement pas le cas: la liste initiale des localités de publication susceptibles d'être extraites du *WoS* n'a pas été constituée de manière structurée. En effet, ce repérage n'est pas le résultat d'une procédure réfléchie, en vue de traitements spatiaux ultérieurs (statistiques, analyse spatiale); c'est simplement le résultat de la saisie de l'adresse fournie par chaque auteur, qui correspond à son lieu d'activité professionnelle.

Le degré de précision spatiale de cette information est variable: certains auteurs indiqueront une ville, parfois un quartier, un campus précis ou un parc technologique, parfois une adresse postale très détaillée. Ce degré de précision dépend aussi des logiques de découpage géographique existant dans chaque pays. C'est ainsi que tous les chercheurs de la région urbaine de Pékin indiquent une adresse (« Pékin »), quand les chercheurs de la région parisienne indiquent le nom de la commune où se trouve leur institution: soit plus de 150 localités « publiantes » dans l'agglomération en 2008 (fig. 2).

Ce degré de précision variable n'est pas sans conséquence sur les résultats et les conclusions tirés de l'exploitation de ces bases de données. D'où la nécessité de procéder à des agrégations spatiales cohérentes au niveau mondial — point traité de manière plus détaillée dans la troisième partie.

Les contenus de ces bases de données « bibliométriques », une fois mis en forme et surtout spatialement homogénéisés, offrent de nombreuses possibilités d'analyse géographique de la production scientifique. On peut privilégier une approche plutôt « verticale » des appartenances territoriales comme les phénomènes d'emboîtement, afin de différencier les espaces étudiés. Il est aussi possible de travailler « horizontalement » sur les interactions entre lieux, les effets de voisinage, le rôle joué par la distance : principalement pour l'étude des coopérations scientifiques (révélées par les cosignatures), mais aussi pour les citations.



Les cinq exemples, pris en Europe, aux États-Unis et en Chine, illustrent les variations de précision de l'information spatiale initiale dans des régions de forte activité scientifique. L'échelle est identique pour les cinq régions urbaines.

2. Publications dans le Web of Science, localisations et zones urbaines, variété des référencements

Ces deux types d'approches resteraient incomplets si la dimension temporelle n'était prise en compte. Observe-t-on par exemple un maintien, une diminution ou, au contraire, une augmentation de la production de tel ou tel groupe de villes, de régions, de pays au cours des deux dernières décennies ? Poser de telles questions suppose que la matrice d'information spatiale rendant compte des publications scientifiques (**encadré**) peut être « répliquable » dans le temps, autrement dit d'année en année. Cela revient à considérer que les dix traitements canoniques de la matrice d'information géographique formalisés par Brian Berry (1964), allant de l'étude des distributions spatiales d'un phénomène à la comparaison d'une sous-matrice au cours du temps en passant par l'analyse des associations ou/et des différenciations spatiales sont toujours possibles. C'est sans doute aller bien vite en besogne car la question de la cohérence temporelle de cette information est aussi importante que celle de la cohérence spatiale.

On ne part bien sûr pas de rien, ce que montre un aperçu rapide de travaux qui ont, au cours des dernières années, utilisé les bases de données bibliographiques à des fins d'analyse géographique de l'activité scientifique.

2. Un bref état de la recherche géographique en matière de production scientifique

L'analyse territoriale de la science est une branche d'un domaine fortement structuré : la « géographie de l'innovation » développée notamment par Maryann Feldman (2002). Cette géographie prend en compte les systèmes d'innovation dans leur dimension territoriale et institutionnelle. Mais certains chercheurs, dont la plus connue est Caroline Wagner (2008) ont formé l'idée que le développement de la « science monde », avec l'augmentation des liens directs entre chercheurs grâce à Internet, permettrait de s'affranchir à la fois de la distance géographique et des ancrages institutionnels et nationaux. La messe serait alors dite : les institutions, universités et organismes de recherche, comme les États ne joueraient plus aucun rôle ; les structures régionales supranationales n'auraient aucune réalité et l'espace géographique ne compterait plus. L'espace de la science serait exclusivement construit par les initiatives des seuls chercheurs. C'est dresser des conclusions hâtives car les travaux de bibliométrie menés par des sociologues, des économistes et des géographes ont produit des résultats qui ne confirment pas ce postulat. On voit, ne serait-ce qu'au niveau des États, que se confirme le maintien des effets d'appartenance, notamment sur la hiérarchie internationale de la production scientifique. Les pays les plus développés (Amérique du Nord, Europe de l'Ouest, Japon) mais aussi les grands pays émergents reconnus comme très dynamiques (Chine, Inde, Brésil, Taiwan, Corée du Sud, etc.) sont les acteurs principaux du monde scientifique (Grossetti *et al.*, 2012). Les résultats présentés par Christian W. Matthiessen sur l'évolution récente des publications dans les principales métropoles mondiales (2010) ne font d'ailleurs sens qu'en prenant en compte les contextes nationaux, cruciaux pour comprendre les dynamiques des pôles scientifiques dans le monde actuel.

2.1. Production scientifique, intégrations régionales et concentrations

Des chercheurs comme Jarno Hoekman, Koen Frenken et Frank van Oort (2009) se sont de leur côté intéressés à l'intégration des espaces scientifiques européens suite aux différentes décisions prises depuis la mise en place des programmes cadres de recherche et de développement (PCRD) au début des années 1980 jusqu'à l'affirmation de la volonté de faire de l'Europe la première économie mondiale de la connaissance. Leurs travaux ont souligné que, si l'intégration européenne progressait entre régions européennes appartenant à des pays différents, cela ne se faisait pas aux dépens des collaborations scientifiques à l'intérieur de chacun des pays. Ils ont également souligné la forte polarisation spatiale de la recherche dans certaines régions, notamment dans les régions capitales, et l'effort fait en direction de ces régions leaders, mettant en évidence la tension entre la politique scientifique de l'Union européenne et sa politique de cohésion territoriale. Cette intégration européenne ne ferait que se superposer à la consolidation des systèmes scientifiques nationaux, conséquence de politiques publiques volontaristes. On souligne aussi l'importance de phénomènes de déconcentration de l'activité scientifique, observables en Russie (Milard, Grossetti, 2006) et dans d'autres pays européens comme la France, l'Espagne : le poids de la capitale y décroît aux profits de centres scientifiques « secondaires » (Grossetti *et al.*, 2009).

Dans les analyses produites, c'est souvent la hiérarchie des espaces de la science qui est étudiée, quand on ne se limite pas à une prise en compte des sommets de cette même hiérarchie. Christian W. Matthiessen et Annette W. Schwarz (1999), travaillant sur l'espace européen, se focalisent ainsi sur les 39 agglomérations d'Europe qui comptaient plus de 5 000 publications en 1994-1996, sans traiter le reste des localités scientifiques. Changeant d'échelle dans un article postérieur (Matthiessen *et al.*, 2002), ils analysent la production des 40 premiers centres mondiaux (période 1997-1999). Dans leur dernier article paru dans *Urban Studies* (2010), ils centrent leur propos sur les 30 à 40 premières métropoles scientifiques, croyant identifier, mais à tort (Grossetti *et al.*, 2012), un poids croissant de ces métropoles dans la production mondiale d'articles.

Ces différentes études ne fournissent qu'une vision partielle de la « science monde » déjà évoquée, ne prenant jamais en compte le système scientifique dans sa globalité et à tous ses niveaux, les résultats sont donc peu assurés. Tout semble donc se passer comme si les analyses des données bibliométriques menées à un niveau géographique plus fin que celui des États conduisaient, pour l'instant, à ne travailler que sur un nombre restreint d'« objets » géographiques. Le défi reste entier : à l'échelle mondiale et au niveau infra-étatique, comment se mettre en situation de mener des recherches sur l'ensemble de la matrice d'information spatiale des productions scientifiques mondiales et pas seulement sur des extractions ?

2.1. Vers une approche urbaine globale

Le niveau géographique de base le plus pertinent pour la constitution d'une information spatiale cohérente semble être, pour des raisons déjà évoquées plus haut, l'échelle urbaine. Le but est donc de constituer une liste mondiale de lieux urbains caractérisés par la présence de l'activité scientifique. Et l'on sent bien qu'une analyse globale de la hiérarchie urbaine scientifique, c'est-à-dire une étude du rôle

des villes dans la structuration spatiale de la science, serait autant d'avancées significatives dans la connaissance de ce phénomène. Souvent les villes ont été approchées par les mailles administratives auxquelles elles appartiennent comme les départements (pour l'Union européenne le niveau dit NUTS 3) ou même les régions (NUTS 2, toujours pour l'Union européenne) (Frenken, 2002; Grossetti, Losego, 2003). Ces « approximations » n'enlèvent rien à la pertinence de certains résultats. En « se contentant » de coder les informations aux niveaux régional ou provincial, Béatrice Milard et Michel Grossetti ont déjà mis en évidence d'importants processus de modification de la carte scientifique en Europe et en Russie (Milard, 2003; Milard, Grossetti, 2006). Si l'on veut aller au-delà de ces « approximations » et mesurer au plus près la part de l'urbain dans les évolutions considérées, on tombe vite sur une difficulté bien connue des géographes : trouver des critères homogènes de définitions de la ville à l'échelle mondiale, qui permettraient des comparaisons internationales de la production et des coopérations scientifiques. Or la science, en tant que système mondialisé de production de connaissances, appelle justement des analyses globales et des comparaisons internationales de l'activité des villes. Christian W. Matthiessen, Annette W. Schwarz et Søren Find ont identifié ce problème (1999, 2002, 2010) et ont finalement choisi de travailler sur des sous-ensembles relativement réduits (30 à 100 métropoles) des villes de la science. D'une certaine façon, ils ont renoncé à caractériser, voire à comprendre, ce qui se passait aux niveaux inférieurs des hiérarchies urbaines. Par exemple, quels peuvent être le poids et le rôle des villes « intermédiaires » dans la production scientifique ? Et que dire de l'émergence des « petites » villes scientifiques ? Quant aux approches « régionales », elles constituent certes une bonne « approximation » quand elles coïncident uniquement avec les plus grandes villes scientifiques. Mais ce n'est pas le cas pour les espaces « régionaux » qui abritent plusieurs villes contribuant de manière équivalente à la production scientifique et dont le fonctionnement peut être plutôt polycentrique ; ni pour les espaces « régionaux » qui abritent des centres de production scientifique d'importance très différente. Dans cette dernière configuration, il convient de cerner non seulement le poids respectif des différentes classes de pôles scientifiques mais aussi d'identifier leurs relations. Tout ceci plaide pour des études, des analyses sur un système urbain dans son ensemble, voire sur les différents systèmes urbains.

Une partie de la solution réside dans un géocodage intégral des lieux de production de la science, puis par la définition de procédures permettant des agrégations spatiales successives : d'abord, au niveau d'une ville ou d'une agglomération, puis éventuellement à des niveaux supérieurs (région urbaine ou administrative, pays, zone transfrontalière, etc.). Ce qui permet d'envisager à la fois une analyse globale du fonctionnement géographique de la science ; et des études multiniveaux. Faut-il privilégier une approche uniquement morphologique pour définir ces agglomérations qui deviennent alors l'élément de base ? Faut-il préférer une approche par les concentrations de populations ou encore combiner ces critères ?

3. Construire une information spatialisée pertinente sur la science : géocodage des adresses et regroupement en entités urbaines

Nous exposons dans cette troisième partie une méthode de codage spatial des données, élaborée dans le cadre du programme ANR Géoscience afin de constituer une matrice d'information géographique des données d'activité scientifique (3). Elle a été développée pour permettre le géocodage intégral de la base Thomson-Reuters et porte sur l'ensemble des localités repérables dans les adresses associées aux articles (4).

Comme indiqué plus haut, les unités spatiales élémentaires de notre travail correspondent aux localités indiquées dans le champ « adresse » des auteurs des articles scientifiques référencés. Cette granularité, très fine pour une étude au niveau mondial, pose deux types de problèmes. Le premier est relatif à la difficulté technique du géocodage : l'opération fondamentale d'attribution d'une localisation géographique précise correspondant aux adresses est complexe. Se pose ensuite la question de l'adéquation des adresses déclarées par les auteurs à notre objectif de localisation et de structuration de l'information en entités géographiques cohérentes. On ne peut les utiliser telles quelles, comme si elles constituaient directement des entités spatiales pertinentes pour l'analyse.

3.1. De l'adresse au géocodage

L'adresse d'un contributeur présente une information géographique dissociée selon plusieurs champs : l'adresse, la ville, le pays et éventuellement la province. On comprend facilement que le codage géographique intégral et manuel de centaines de milliers d'adresses n'est pas réalisable (tableau 1). Fort heureusement, les informations contenues dans l'adresse peuvent permettre l'utilisation d'outils de géocodage automatique, disponibles en ligne, qui affectent une latitude et une longitude aux localités à partir de leurs bases de données de référence. Mais l'affaire n'est pas si simple. Il ne suffit pas de laisser tourner un utilitaire de géocodage pour avoir une affectation géographique fiable de ces données. Il faut procéder à un nettoyage préalable des adresses, ce qui s'avère être une correction plus ou moins complexe suivant les cas de figure. La seule correction de l'orthographe des noms de lieux est longue et minutieuse. Pour de nombreuses raisons (graphies multiples, erreurs de saisie des auteurs ou des opérateurs de la base, translittération) les lieux de publication apparaissent sous de multiples orthographes. Par exemple, le toponyme Tao-Yuan a 10 variantes d'orthographe dans la base WoS, et correspond en fait à trois villes différentes, la principale à Taiwan, les deux autres en Chine (5). Par ailleurs, la dimension historique est aussi d'une grande importance, les villes, les provinces, les pays changeant de nom ou d'affectation au fil du temps : Leningrad redevient Sankt-Peterburg, généralement orthographié Saint-Petersburg en anglais ; l'Allemagne est réunifiée et Berlin redevient une seule ville, Hong-Kong est rendue à la Chine, etc.). Enfin, ces trois champs (ville, province, pays) sont parfois insuffisants pour déterminer une localisation précise. En effet, l'information sur la province est souvent absente de la base, elle n'est le plus souvent indiquée que pour les pays présentant une organisation fédérale, avec un premier niveau administratif très marqué, comme les USA, le Canada, le Brésil, l'Inde, etc. Ce qui signifie que sont alors confondues les localités homonymes à l'intérieur d'un même pays, problème non

négligeable par exemple en France dans le cas des toponymes issus de saints patrons (Saint-Denis). Dans certains cas, un retour à la source des données a été nécessaire pour pouvoir répartir la production scientifique entre les différentes localités homonymes.

1. Année bibliographique 2008 : le travail de codage des adresses du <i>Web of Science</i>	
Information source (nombre de notices d'articles)	2 074 000
Localités associées à ces notices (avant correction)	31 680
Localités identifiées (corrigées et géocodées)	18 300

Comment, une fois ces corrections et harmonisations effectuées, mettre en place une procédure de géocodage automatique ? On ne pouvait se contenter de choisir une solution « clés en main ». En effet, travailler sur des informations à l'échelle mondiale et sur une période d'une quarantaine d'années élimine les outils de géocodage commerciaux les plus courants, adaptés à un pays précis et à des toponymes actuels. Les autres solutions accessibles se regroupent en deux catégories : les bases de données de toponymes et les outils web de géocodage automatique. Les premières, comme par exemple la base GeoNames (6), malgré leur extension et la gestion de toponymes variés pour un même lieu, se sont révélées insuffisantes, le codage géographique étant inégal et hétérogène géographiquement. Le second type d'outils, basé sur des services web, s'est révélé plus performant, mais sa mise en œuvre plus complexe techniquement (7). Au final, Le service le plus performant pour géocoder nos localités a été celui de Google (8), qui de plus informe sur la qualité des résultats fournis. Le résultat n'est toutefois pas exempt d'erreurs : certaines confusions de noms de quartiers, de villes et de noms de provinces n'ont pu être évitées. Comment vérifier ces données ? Question difficile, car le contrôle de cohérence avec des données plus ou moins équivalentes est impossible (9). Cette phase de vérification et de correction des données (2011-2012) a été fondamentale pour l'amélioration des résultats du géocodage « brut » issus du service Google. Pour ne donner que deux exemples, l'outil, traitant le nom « Garching », avait attribué à un obscur hameau de Bavière le statut de pôle scientifique, alors que cette adresse renvoie en fait à un parc technoscientifique de la banlieue de Munich, où sont rédigées des centaines d'articles scientifiques chaque année. Il en va de même pour une localité proche de Chicago où il n'y a pas de centre de recherche, mais qui est l'homonyme d'un « véritable » et important pôle situé, lui, dans l'agglomération (Argonne National Laboratory). On s'intéressait, dans ces deux, cas aux anomalies « visuelles » apparaissant dans notre outil de vérification cartographique en ligne ; dans d'autres cas, ce sont les valeurs plutôt surprenantes de certaines villes qui pouvaient susciter le doute et déclencher des vérifications. Enfin, un retour vers le nombre d'articles pris en compte par notre procédure de géocodage a été indispensable. Il s'agissait de savoir si les adresses géocodées pouvaient être associées à un pourcentage satisfaisant de la production scientifique de chaque pays, ou de chaque région d'un pays quand l'information désagrégée au niveau régional était disponible. On obtenait ainsi une forme de « taux de couverture » du contenu de la base et donc une évaluation de la qualité du géocodage dans chaque pays.

Une fois les principales phases de contrôle terminées, on obtient des données souvent localisées de manière fine (France, États-Unis, [fig. 2](#)). Car la précision de la localisation, notamment intra-urbaine, demeure variable suivant les pays; les données ne sont donc pas strictement comparables. Par ailleurs, le niveau maximal de précision obtenu avec le géocodage (sa granularité) n'a pas forcément de sens pour analyser la répartition spatiale mondiale des lieux de production de la science. Pour nous, les lieux pertinents sont plutôt des pôles urbains ou des agglomérations. Enfin, la précision de certaines adresses à l'échelle intra-urbaine est parfois illusoire : un même institut de recherche peut, dans une ville, avoir des implantations très différentes et l'adresse déclarée par un chercheur renvoyer au centre administratif de l'institut. Il a donc fallu proposer une méthode d'agrégation de nos données de base en entités géographiques (villes, agglomérations) plus cohérentes.

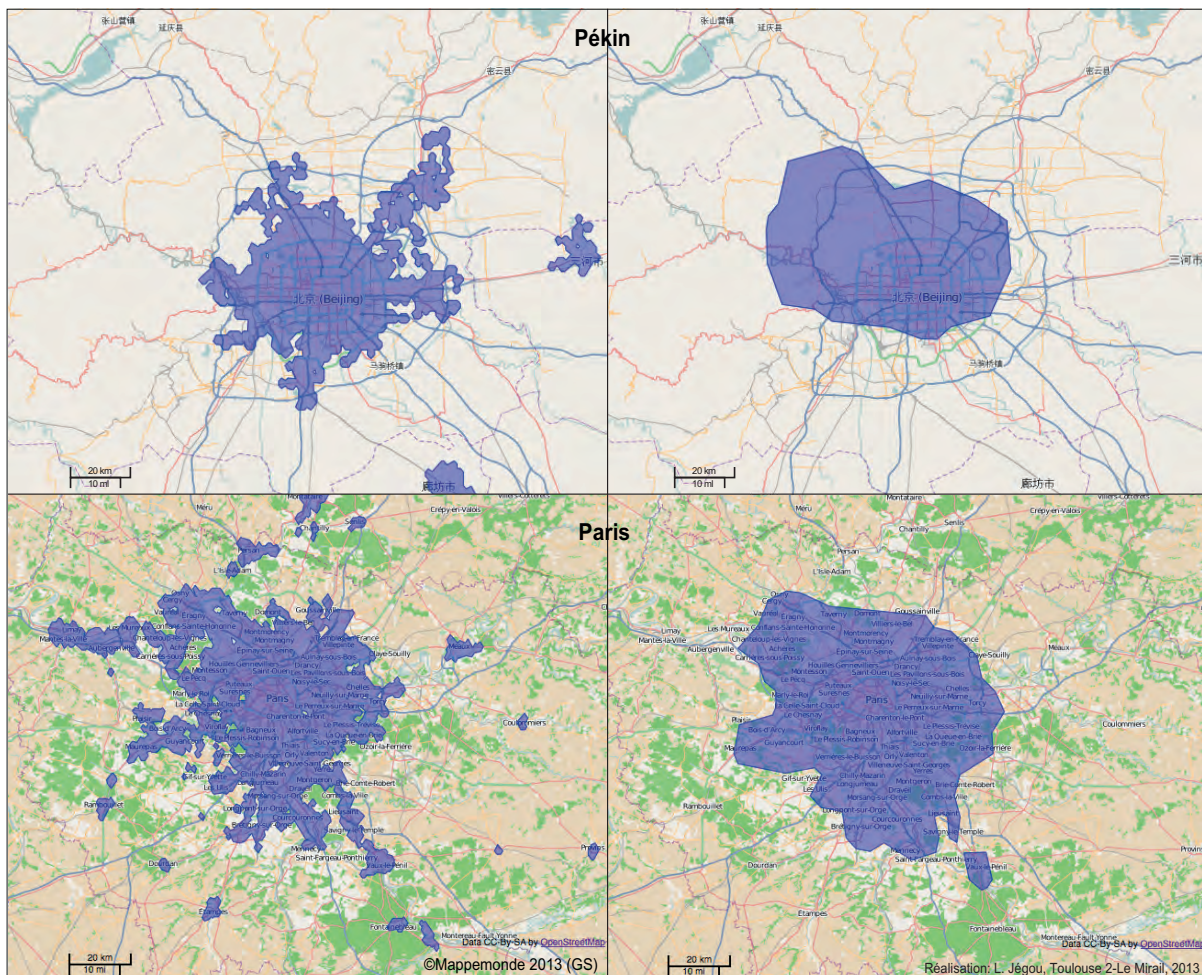
3.2. Du lieu de publication à l'agglomération

Une fois établi le caractère hétérogène et peu utilisable de l'information spatialisée à son niveau de base (l'ensemble des adresses géocodées), il faut déterminer un critère de regroupement de ces localités en agglomérations plus pertinentes, comparables entre elles. Les études précédentes sur la géographie de la production scientifique ont utilisé des systèmes de regroupement classiques, fondés sur les niveaux administratifs et leur assemblage en groupes urbains (Pumain, 2003). Cette méthode est dépendante de la maille de ces découpages administratifs, qui est très largement variable de par le monde, et ne peut s'appliquer aux grandes conurbations mondiales, car elle présuppose la ville comme une entité incluse dans une région administrative qui l'englobe. Notre démarche a été de privilégier le critère de l'espace « scientifique urbain », en s'intéressant à la fois au caractère urbain au sens morphologique et de densité de population, mais aussi au critère d'une certaine densité d'activité scientifique, qui ne se superpose (spatialement) pas forcément avec le critère précédent. Cette proposition n'épuise pas toutes les possibilités. Les données une fois géocodées se prêtent à de multiples méthodes d'agrégation en fonction des besoins et des échelles d'analyse envisagées.

Pour juger du caractère urbain de la localisation d'une publication scientifique, il nous fallait avoir accès à une définition mondiale et homogène des zones urbaines, associée à des délimitations cartographiques précises. Il existe plusieurs sources d'information sur les zones dont l'occupation du sol est qualifiée d'urbaine, issues du traitement d'images satellites. La détermination du caractère « urbain » est dans ce cas réalisée essentiellement à partir du critère d'anthropisation de l'occupation du sol et de sa morphologie. Deux sources de ce type ont été prises en compte : le programme Ionia GlobCover de l'ESA (10) et le programme Global Urban Extent dirigé par A. Schneider et M. Friedl (2009, 2010). Les données produites par ces sources présentent l'avantage d'être homogènes sur le globe et de se fonder sur des critères assez objectifs (voire automatiques) d'anthropisation des sols. Cependant, la ville ne consiste pas seulement en des sols d'aspect non « naturel », et la question de la délimitation des grandes conurbations demeure. L'autre solution classique pour délimiter la ville revient à utiliser les densités de population, qui constituent très souvent un des critères mêmes de définition de la ville. Là aussi, il existe des sources de données sérieuses et de couverture mondiale, proposant des valeurs localisées de densité pour plusieurs dates. Nous avons utilisé les données du programme

Global Population of the World, du laboratoire SEDAC à l'Université de Columbia (11). La densité variant fortement dans le monde, et, par suite, le niveau au-delà duquel on se considère « en ville », un simple seuillage des données de densité ne suffisait pas. Il a donc fallu trouver un indicateur permettant de déterminer spatialement les fortes variations de cette densité, dans l'espoir de produire une délimitation utilisable. C'est justement la fonction des LISA (Local Indicators of Spatial Association), dont le I local de Moran qui indique de façon homogène sur un territoire les regroupements statistiquement significatifs ou « noyaux de densité » (Anselin, 1995). C'est une méthode manuelle assistée qui a finalement été utilisée pour regrouper les localités publiantes en agglomérations. Le critère essentiel a été la zone de densité homogène au sens d'Anselin, parfois améliorée en fonction du résultat de la visualisation simultanée de toutes les méthodes grâce à notre outil de *webmapping* interactif.

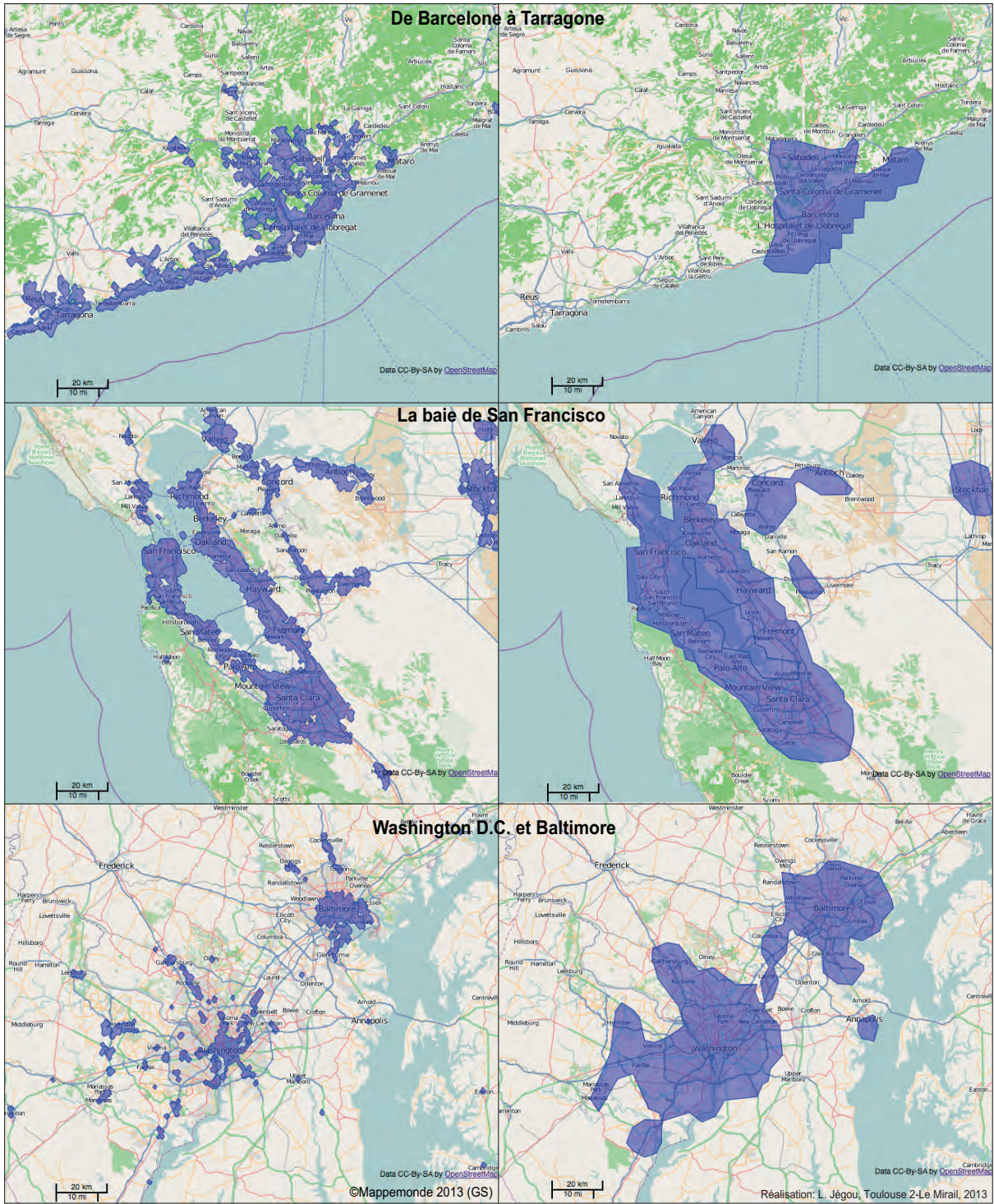
On voit bien comment la délimitation par noyaux de densité permet plus facilement de constituer des ensembles relativement homogènes, plus appropriés à notre démarche (fig. 3a et 3b).



Sélection des occupations « urbaines » du sol

Forte variation de la densité de population

3a. Critères de délimitation des zones urbaines, variations des résultats



Sélection des occupations « urbaines » du sol Forte variation de la densité de population
3b. Critères de délimitation des zones urbaines, variations des résultats

On aboutit à la création d'agglomérations *ad hoc*: les critères de regroupement sont constants pour l'ensemble des grandes zones urbaines du globe. Les lieux de publication(s) situés à plus de 40 km d'un de ces regroupements ont fait l'objet d'un traitement particulier (12).

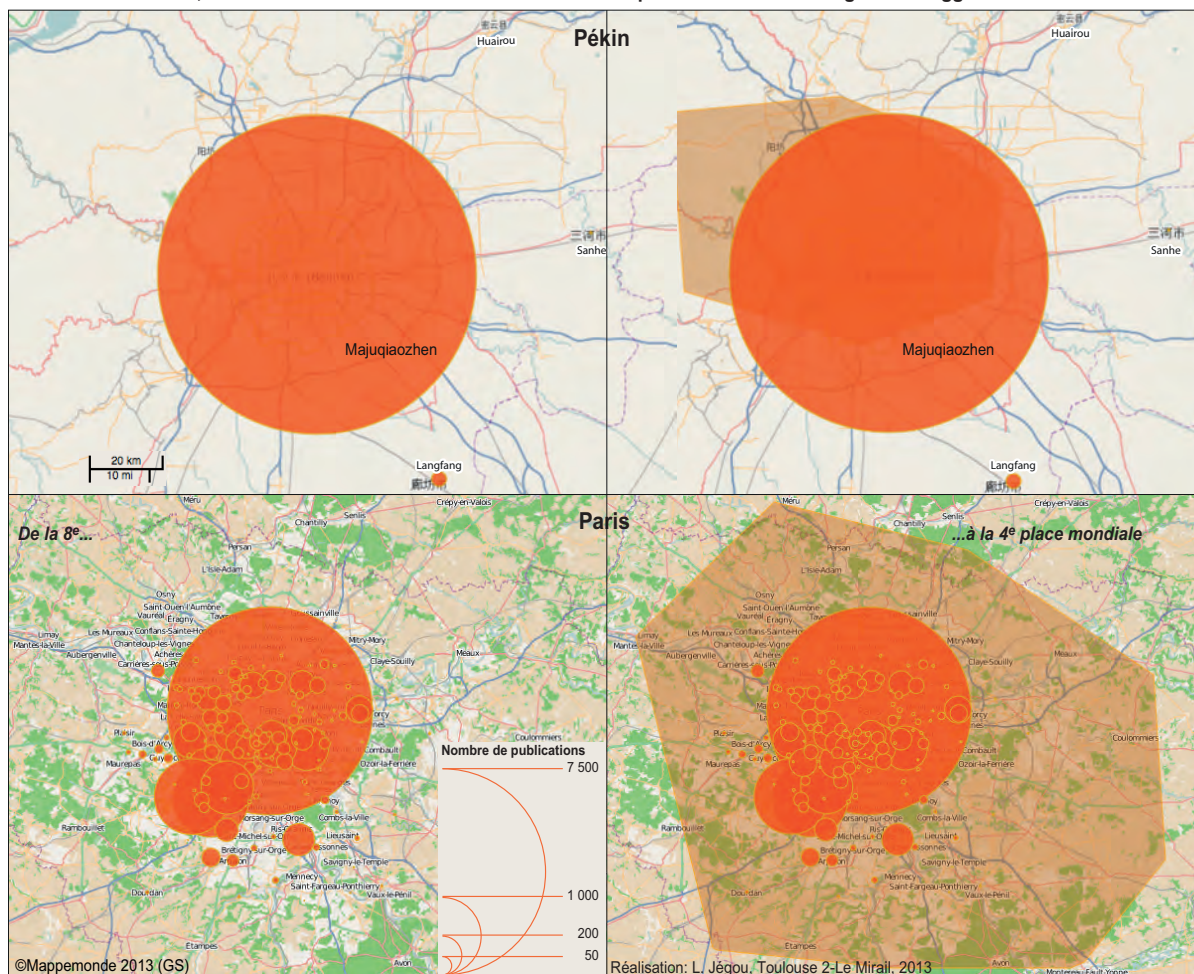
On voit comment (fig. 4a et 4b) on arrive à la constitution, dans cinq exemples, d'agglomérations « scientifiques » et au regroupement des publications qui y ont été localisées. L'agglomération « scientifique » de Barcelone englobe la proche banlieue, mais pas la ville proche de Tarragone alors qu'il existe une continuité morphologique (mais une rupture de densité de population). À l'inverse, on arrive à un large regroupement des pôles scientifiques de la baie de San Francisco en une unité, de même que les alentours de Baltimore et Washington sont considérablement simplifiés en fonction de ces deux entités principales qui agrègent l'essentiel de l'activité scientifique du Sud de la Megalopolis.

Les exemples choisis soulignent bien l'effet que de tels regroupements peuvent avoir sur la position d'une ville ou d'un pôle scientifique dans l'ensemble mondial du système des publications scientifiques. Le rang d'une ville est extrêmement sensible à la manière dont on agrège, ou non, les données source. La région urbaine de San Francisco et celle de Paris en sont les meilleurs exemples. La fragmentation administrative de la région parisienne (fig. 4a) ferait presque oublier qu'elle constitue en fait le 4^e pôle mondial (données WoS, 2008). Autour de la baie de San Francisco (fig. 4b), la répartition très polycentrique des centres de recherche fait que seule l'agrégation spatiale rend possible une évaluation pertinente de son poids : c'est le 10^e pôle mondial en 2008, alors que la seule ville de San Francisco émergerait à peu près au 60^e rang mondial, sans Palo Alto, Berkeley, Stanford. Nous avons pris, dans cette figure 4, l'évolution de la position de quelques villes bien connues comme exemple de la sensibilité des données à l'agrégation spatiale. Notre objectif scientifique n'est cependant pas de construire des classements, palmarès et autres « rankings » de la science qui fleurissent et sont autant d'obstacles à la compréhension de la structuration de la production scientifique, focalisant le regard sur quelques entités géographiques du système mondial (les « 10 » ou « 30 » premières villes par exemple). Au contraire, nous souhaitons pouvoir analyser les modalités du fonctionnement territorial global de la science en tentant une approche multiniveau, analyser les interactions entre tous les lieux, mesurer le poids respectif des centres petits, moyens ou grands, etc. L'une de nos premières analyses issues de ce travail de construction géographique nous a d'ailleurs amenés à mettre en évidence un puissant mouvement de diffusion et donc de déconcentration spatiale de la science mondiale, qui rend d'autant plus vaine la focalisation vers les quelques dizaines de très grandes métropoles dont l'importance dans le système mondial est très souvent surestimée (Grossetti *et al.*, 2012).

Notre démarche a aussi une caractéristique : les agrégations spatiales, réalisées dans les couches de notre modeste « SIG de la science » sont transformables, réformables et réévaluables. Elles n'ont pas vocation à être figées. Les données de base sont réassociables en fonction de problématiques et d'échelle d'analyse spécifiques, ainsi que des progrès faits dans la compréhension des logiques d'organisation spatiale de la science.

Publications du WOS, avant ...

... après délimitation homogène des agglomérations



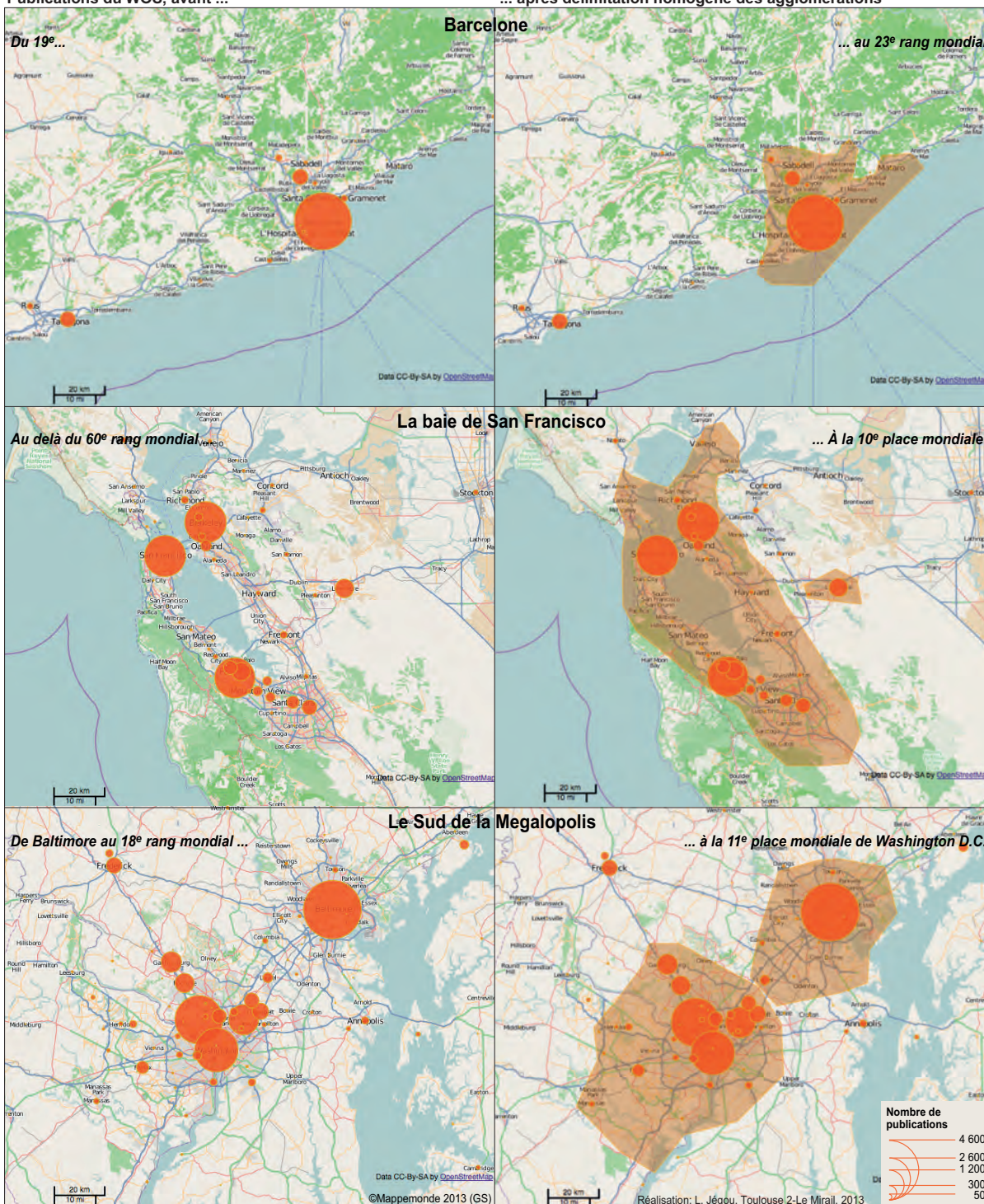
Publication du Web of Science avant...

... après délimitation homogène des agglomérations

4a. Publications dans le Web of Science et hiérarchie urbaine mondiale

Publications du WOS, avant ...

... après délimitation homogène des agglomérations



Publication du Web of Science avant...

...après. délimitation homogène des agglomérations

4b. Définition homogène des agglomérations, publications dans le Web of Science et hiérarchie mondiale

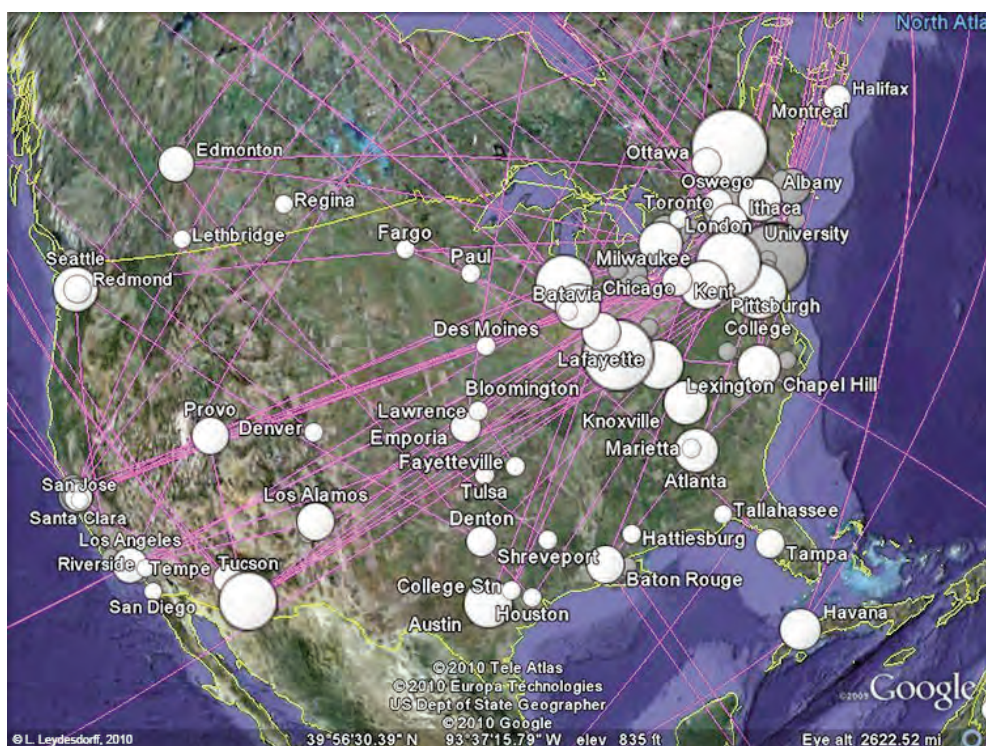
Conclusion : Places et apports des représentations cartographiques de la production scientifique

L'analyse géographique et la mise en carte des données relatives aux productions scientifiques se jouent sur deux scènes : d'une part, celle de l'exploration cartographique après codage minutieux, qui correspond au travail sur les données elles-mêmes, d'autre part, celle de la sélection des résultats pertinents et de leur exposition ordonnée, qui correspond à la structuration et l'agrégation des entités spatiales de base.

À notre connaissance, aucun géocodage intégral des données scientifiques n'a été proposé en prenant en compte l'ensemble de ces aspects. Certaines visualisations de la « science mondiale » ont été réalisées. Les « cartes » de L. Leydesdorff qui ont abondamment circulé et sont consultables sur son site (13) (2010) sont assez inadaptées au phénomène à représenter, le géocodage étant loin d'être exhaustif et la réflexion sur l'agrégation spatiale absente. Pour s'en convaincre, il suffit de mettre en regard la cartographie des publications américaines en science de l'information et de la communication proposée par Leydesdorff (fig. 5) et celles résultant, d'une part, de notre géocodage primaire (fig. 6a) et, d'autre part, des regroupements d'entités spatiales comparables (fig. 6b) (14).

Lorsque des chercheurs s'attaquent à l'autre versant du problème : l'agrégation en unités spatiales pertinentes (et notamment en agglomérations) des données d'activité scientifique, ils se restreignent jusqu'ici, on l'a vu, à des sous-ensembles de la population des villes mondiales. Notre proposition est, à ce jour, la seule qui prenne en compte l'ensemble du périmètre du *Web of Science*, en faisant des propositions pour rendre les unités spatiales de bases comparables (Charre, 1995).

Notre travail, qui vise à mettre en adéquation les contenus du *WoS* avec les prescriptions de Brian Berry énoncées voici presque cinquante ans, peut finalement permettre de produire certes de nouvelles « cartes de la science » reflétant les distributions spatiales de l'activité de publication des chercheurs dans l'ensemble du monde, pour les années où la couverture géographique de la base peut être considérée comme correcte. Mais l'objectif est aussi de structurer les données d'activité scientifique d'un point de vue géographique, afin de permettre des traitements non-cartographiques : proposer des clés d'agrégation spatiale modulables en fonction du niveau d'analyse autorise le développement d'une réflexion sur la structure des réseaux de collaboration et l'élaboration de graphes. Le travail de mise en cohérence spatiale de ces données permet finalement de produire des tableaux et des matrices d'information spatiale qui font davantage sens que ceux qui résultent d'une extraction brute des données localisables du *WoS* ou de *Scopus*.



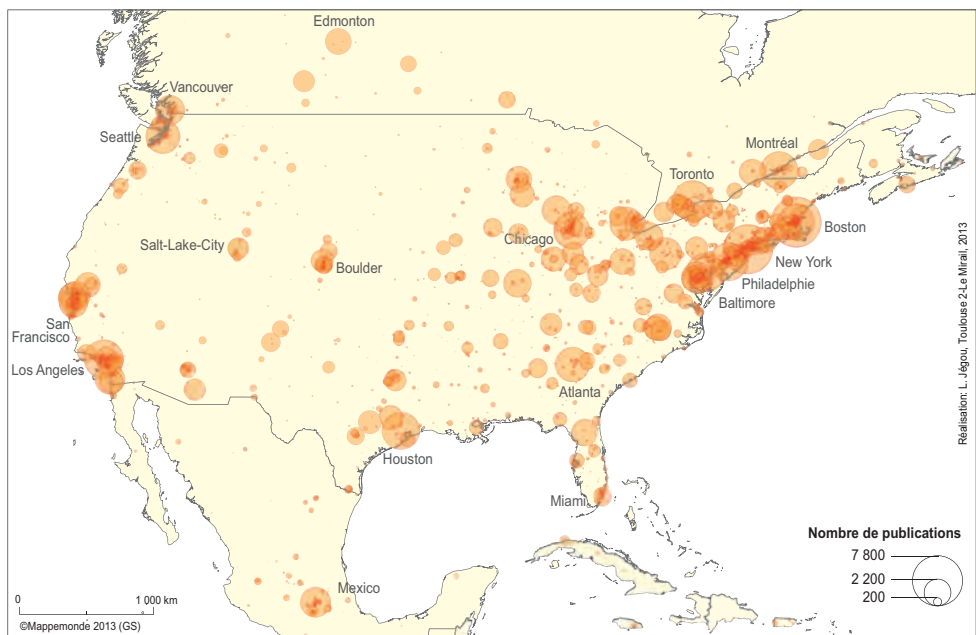
5. Une géovisualisation récente des données de production scientifique, via Google Earth. Avec l'aimable autorisation de l'auteur (© L. Leydesdorff).

Cette géovisualisation est l'un des premiers exemples de cartographie des bases de données bibliométriques (ici, le WoS) à échelle fine. Elle montre les lieux essentiels d'où provenaient les articles publiés dans le domaine des « Library and Information Sciences » en 2009. Une version en a été publiée en 2010 dans *JASIST* (Leydesdorff, Persson, 2010). Loet Leydesdorff a mis au point une procédure fondée sur une série de logiciels, décrite précisément dans l'article en question, pour géocoder les articles avec l'outil automatique fourni par Yahoo!, puis générer un fichier KML dont la visualisation est possible avec le logiciel Google Earth. Les données sources n'ont pas été agrégées par zone urbaine (on a par exemple, dans la région de San Francisco, plusieurs cercles superposés : Santa Clara, San José...) et sont représentées au niveau des localités scientifiques « primaires ».

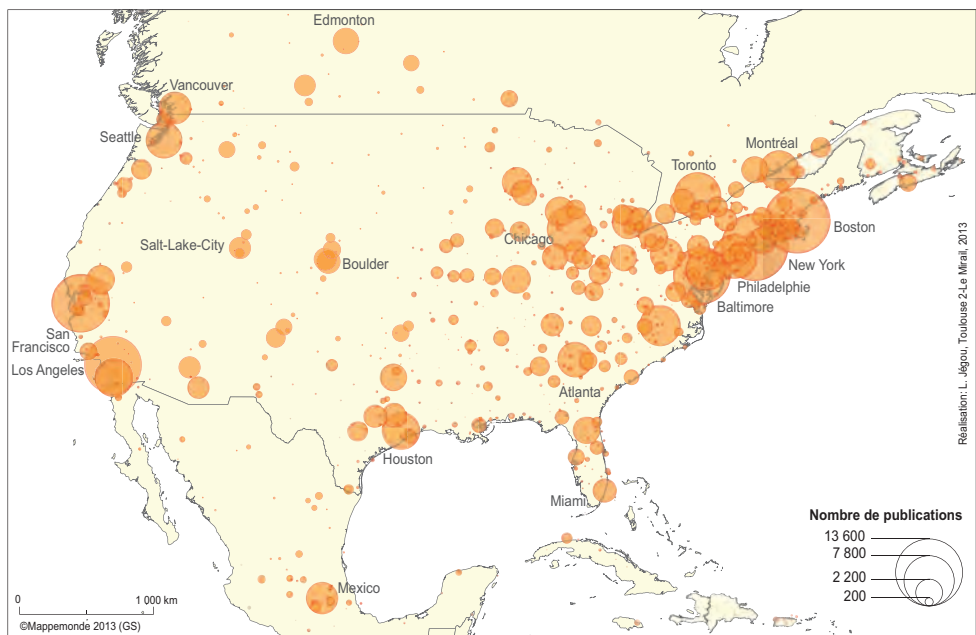
On voit ici l'intérêt des logiciels de géovisualisation pour l'exploration des données par le chercheur. Il faut souligner d'ailleurs le caractère pionnier du travail de Loet Leydesdorff dans l'élaboration de cartes de la science.

Mais apparaissent aussi les limites des utilitaires de géovisualisation quand il s'agit d'aller vers une représentation cartographique de qualité. Ici Google Earth impose à l'utilisateur un fond graphiquement surchargé et aux informations inutiles (relief, bathymétrie...), avec des couleurs criardes aux limites de la lisibilité. La manière dont le logiciel gère l'affichage des noms complique encore la donne. Enfin, Google Earth ne crée pas de légende et ne permet pas de hiérarchiser le dessin des symboles pour éviter les superpositions.

À l'appui de cet exemple, on regrette que la qualité médiocre des utilitaires de géovisualisation aujourd'hui ne serve pas mieux les chercheurs qui entreprennent, par des moyens facilement accessibles, de mettre en cartes leurs données. On souhaite vivement la popularisation de solutions logicielles simples permettant à une plus large communauté d'utilisateurs une cartographie plus efficace et épurée. Mais c'est là un débat qui dépasse largement le cadre de la géographie de la science et dont *M@ppemonde* s'est d'ailleurs plusieurs fois fait l'écho.



6a. Les lieux de publications initiaux. Localités et nombre de publications en 2008. Cette carte représente, pour la seule Amérique du Nord, l'ensemble des lieux de publications présents dans notre base de données. Dans les zones où l'activité scientifique est forte, avec de multiples points situés dans la même agglomération, l'importance des concentrations métropolitaines est masquée. D'où un « biais visuel » qu'il convient de corriger. Cette carte n'est donc, de notre point de vue, qu'un document intermédiaire dans la représentation, à cette échelle, des territoires de la science.



6b. Les publications agrégées par agglomérations. Localités agglomérées et nombre de publications en 2008. Les lieux de publication ont été regroupés par agglomérations. Si la carte reste visuellement complexe, le rapport entre petites villes et grandes agglomérations est restitué avec plus de justesse. On voit beaucoup mieux l'importance fondamentale des villes de la mégapole comme des grandes conurbations californiennes, sans que ne disparaisse une autre information: le maillage assez serré du territoire étatsunien par des lieux de production scientifique, témoin avant tout de la répartition des centres universitaires dans l'ensemble du pays.

6. Une cartographie systématique des publications

Bibliographie

- ANSELIN L. (1995). « Local indicators of spatial association – LISA ». *Geographical Analysis*, vol. 27, n° 2, p. 93-115. doi:10.1111/j.1538-4632.1995.tb00338.x
- BARON M., ECKERT D., JÉGOU L. (2011). « Peut-on démêler l'écheveau mondial des coopérations scientifiques ? ». *M@ppemonde*, n° 102. <http://mappemonde.mgm.fr/num30/internet/int11201.html>
- BERROIR S., CATTAN N., SAINT-JULIEN Th. (2009). « Les masters en réseau : vers de nouvelles territorialités de l'enseignement supérieur en France ». *L'espace géographique*, vol. 38, n° 1, p. 43-58. http://www.cairn.info/load_pdf.php?ID_ARTICLE=EG_381_0043
- BERRY B.J.L. (1964). « Approaches to Regional Analysis : A Synthesis ». *Annals of the Association of American Geographers*, vol. 54, n° 1, p. 2-11. doi:10.1111/j.1467-8306.1964.tb00469.x
- BRUNET R. (1990). « L'université, la ville et la région ». *L'Espace géographique*, vol. 19-20, n° 3, p. 212-213.
- CHARRE J. (1995). *Statistique et territoire*. Montpellier : Reclus, coll. « Espaces modes d'emploi », 119 p. ISBN : 2-86912-060-5
- FELDMAN M. P., MASSARD N., dir. (2002). *Institutions and Systems in the Geography of Innovation*. Boston, Dordrecht, London : Kluwer Academic Publishers, coll. « Economics of science, technology and innovation », 368 p. ISBN : 0-7923-7614-5
- FRÉMONT A. (1990). « L'aménagement du territoire universitaire français ». *L'Espace géographique*, vol. 19-20, n° 3, p. 193-202.
- FRÉMONT A., HÉRIN R., JOLY J., dir. (1992). *Atlas de la France universitaire*. Montpellier : GIP Reclus ; Paris : la Documentation française, coll. « Dynamiques du territoire », 270 p. ISBN : 2-11-002732-0
- FRENKEN K. (2002). « A New Indicator of European Integration and an Application to Collaboration in Scientific Research ». *Economic Systems Research*, vol. 14, n° 4, p. 345-361. doi:10.1080/0953531022000024833
- FRENKEN K. (2009). « Spatial scientometrics : Towards a cumulative research program ». *Journal of Informetrics*, n° 3, p. 222-232. <http://www.sciencedirect.com/science/article/pii/S1751157709000303>. doi:10.1016/j.joi.2009.03.005
- GROSSETTI M., ECKERT D., GINGRAS Y., JÉGOU L., LARIVIÈRE V., MILARD B. (2012), « The Geographical Deconcentration of Scientific Activity (1987-2007) », *Proceedings of the 17th Conference on Science and Technology Indicators*, Montréal, p. 348-356. http://sticonference.org/Proceedings/vol1/Grossetti_Geographical_348.pdf
- GROSSETTI M., LOSEGO Ph., MILARD B. (2009). « La territorialisation comme contrepoint à l'internationalisation des activités scientifiques ». In LAREDO P., LERESCHE J.-Ph., WEBER K., *L'internationalisation des systèmes de recherche en action. Les cas français et suisse*. p. 281-300. <http://halshs.archives-ouvertes.fr/halshs-00471192> oai:halshs.archives-ouvertes.fr:halshs-00471192
- GROSSETTI M., LOSEGO P., dir. (2003). *La Territorialisation de l'enseignement supérieur et de la recherche*. Paris, Turin, Budapest : L'Harmattan, coll. « Géographies en liberté », 339 p. ISBN : 2-7475-4283-1

- HOEKMAN J., FRENKEN, K., VAN OORT F. (2009). «The geography of collaborative knowledge production In Europe». *The Annals of Regional Science*, vol. 43, n° 3, p. 721-738. doi:10.1007/s00168-008-0252-9
- LANE D., PUMAIN D., VAN DER LEEUW S., WEST G., dir. (2009). *Complexity perspectives on innovation and social change*. Dordrecht: Springer, ISCOM, coll. «Methodos». ISBN: 978-1-4020-9662-4
- LARGERON C., MASSARD N. (2001). «La géographie des collaborations scientifiques en France: une étude de la structuration des co-publications entre départements français». *Revue d'économie régionale et urbaine*, n° 1, p. 39-52. doi:10.3917/reru.011.0039
- LATOUR B., WOOLGAR S. (1979). *Laboratory life. The social construction of scientific facts*. Beverly Hills (Californie): Sage Publications, 272 p. ISBN: 0-8039-0993-4
- LEYDESDORFF L., PERSSON O. (2010). «Mapping the Geography of Science: Distribution Patterns and Networks of Relations among Cities and Institutes». *Journal of the American Society for Information Science & Technology*, vol. 61, n° 8, p. 1622-1634. doi:10.1002/asi.21347
- MASSARD N., AUTANT-BERNARD C., RIOU S. (2003). *Production de connaissances et innovation: positionnement et enjeux pour la MIAT Bassin parisien*. Saint-Étienne: Convention Creuset/MIAT Bassin parisien, rapport final, étude «innovation et développement local», 118 p.
- 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*, vol. 47, n° 9, p. 1879-1897. doi:10.1177/0042098010372683
- 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 Bibliometrics Indicators». *Urban Studies*, vol. 39, n° 5-6, p. 903-927. doi:10.1080/00420980220128372
- MATTHIESSEN C.W., SCHWARZ A.W. (1999). «Scientific Centres in Europe: An Analysis of Research Strength and Patterns of Specialisation Based on Bibliometric Indicators». *Urban Studies*, vol. 36, n° 3, p. 453-477. doi:10.1080/0042098993475
- MILARD B. (2003). «Territorialisation de la production scientifique dans le sud-ouest européen et Collaborations scientifiques et territoires dans le sud-ouest européen». In GROSSETTI M., LOSEGO Ph., dir., *La Territorialisation de l'enseignement supérieur et de la recherche en Europe. France, Espagne, Portugal*. Paris: l'Harmattan, coll. «Géographies en liberté», 339 p. ISBN : 2-7475-4283-1
- MILARD B., GROSSETTI M. (2006). «L'évolution de la recherche scientifique dans les régions de Russie: déclin ou déconcentration? ». *M@ppemonde*, n° 81, 13 p. <http://mappemonde.mgm.fr/num9/articles/art06105.html>
- PUMAIN D. (2003). «Scaling laws and urban systems». Working paper, SFI Working Group on Social Scaling, Santa Fe Institute. <http://www.santafe.edu/media/workingpapers/04-02-002.pdf>
- SCHNEIDER A., FRIEDL M. A., POTERE D. (2009). «A new map of global urban extent from MODIS data». *Environmental Research Letters*, vol. 4, article 044003.

- SCHNEIDER A., FRIEDL M. A., POTERE D. (2010). « Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions' ». *Remote Sensing of Environment*, vol. 114, n° 3, p. 1733-1746. doi: 10.1016/j.rse.2010.03.003
- VASSAL S. (1969). « Les nouveaux ensembles universitaires français. Éléments de géographie urbaine ». *Annales de géographie*, vol. 78, n° 426, p. 131-157. doi: 10.3406/geo.1969.15836
- WAGNER C. (2008). *The New Invisible College: Science for Development*, Washington: Brookings Institution Press, 175 p. ISBN: 978-0-8157-9213-0
- WOUTERS P., DUVAL J. (2006). « Aux origines de la Scientométrie. La naissance du Science Citation Index ». *Actes de la Recherche en Sciences Sociales*, n° 164, p. 11-22. doi:10.3917/arss.164.0011

Notes

1. D'une discipline à l'autre, d'une communauté nationale à une autre, la publication dans une revue internationale n'a pas forcément le même sens ou la même fréquence. Les mathématiques n'ont évidemment pas les mêmes pratiques que des sciences appliquées comme par exemple le génie chimique. Il n'en reste pas moins que ces bases, qu'il s'agisse du *Web of Science* ou de *Scopus*, constituent des outils incontournables pour appréhender la production scientifique mondiale.
2. *WoS*: ensemble de bases de données bibliographiques maintenu par Thomson-Reuters, et qui comprend notamment le fameux *Science Citation Index* (SCI). *Scopus*: ensemble équivalent de bases de données bibliographiques, maintenu par l'éditeur Elsevier.
3. Le géocodage et sa vérification ont été réalisés avec l'aide de Fabien Goblet.
4. La mise en forme des données, géocodage, définition des périmètres d'agglomération et vérification/correction des données, a pris environ 18 mois, impliquant de nombreux participants au programme ANR.
5. Tao-Yuen, Tasyuan, Tao-Yuan, Taoyuan-County, Taoyuan, Tauyuan, Taoyuan-Cty, Taoyuang, Tau-Yuan, Taoyuan-Hsien, Tayuan
6. <http://www.geonames.org/>
7. Il faut programmer des scripts spécifiques pour interroger ces services et vérifier les résultats.
8. Google Geocoding API: <https://code.google.com/apis/maps/documentation/geocoding/>, cependant limité à 2 500 adresses par jour.
9. Pour ce faire, nous avons réalisé un outil en ligne de visualisation cartographique des données de publication, interne au groupe Géoscience. On a donc décidé de répartir les vérifications entre différents chercheurs du projet Géoscience, en fonction de leurs connaissances du système scientifique et de la géographie de tel ou tel pays.
10. <http://ionia1.esrin.esa.int>, ESA GlobCover Project, MEDIAS-France et UCLouvain
11. <http://sedac.ciesin.columbia.edu/gpw/>
12. Ils ont été laissés seuls s'ils étaient réellement isolés, c'est à dire à plus de 40 km de n'importe quel autre point isolé. Dans le cas contraire, ils ont été regroupés avec leur voisin le plus « publiant » : ce qui permet par exemple d'associer les villes très proches de Homburg et Saarbrücken (Allemagne) ou Saint Andrews et Dundee (Ecosse) pour former un agrégat suffisamment éloigné des grandes agglomérations. Notre postulat est que cette grande proximité géographique (certainement moins d'une heure de transport dans tous les cas de figure) fait sens.
13. <http://www.leydesdorff.net/index.htm>

14. Dans le même registre, on rappellera la tentative de (pseudo)cartographie des collaborations scientifiques mise en ligne par Olivier Beauchesne au début de l'année 2011. Il s'agissait uniquement d'une représentation sans traitement préalable et sans commentaire, sans tentative de valuer les liens figurés. Cette image cartographique assimilait de manière très contestable les cosignatures d'articles aux amitiés Facebook (Baron *et al.*, 2011).

Adresses des auteurs

Denis Eckert, UMR LISST, Université Toulouse 2-Le Mirail, 5 allées Antonio Machado, 31058 Toulouse Cedex 9. Courriel : eckert@univ-tlse2.fr

Myriam Baron, Géographie-Cités, Équipe P.A.R.I.S., Université Paris Diderot-Paris 7-UFR GHSS 75205 Paris Cedex 13. Courriel : baron@parisgeo.cnrs.fr

Laurent Jégou, département de Géographie-Aménagement-Environnement, Université Toulouse 2-Le Mirail, 5 allées Antonio Machado, 31058 Toulouse Cedex 9. Courriel : jegou@univ-tlse2.fr