Komaba Studies in Human Geography
Vol.24 45-60 2022
Advance Online Publication

Constraints and Opportunities in Mapping Japanese Patent Information

BAEZA-GONZALEZ Sebastian* and KAMAKURA Natsuki**

*Departamento de Geografía, Facultad de Arquitectura, Urbanismo y Geografía, Universidad de Concepción.
**Collaborative Research Organization for Future Regional Society, the University of Tokyo.

Abstract Economic geographers and innovation studies, in general, have paid much attention to the formation, distribution, and dynamics of knowledge in space for decades. Some of this analysis uses patent information as a proxy to understand these issues, particularly as a source for spatializing knowledge creation and innovation. In this article, we use patent information to identify clusters in different technology domains in Japan. We conduct this exercise to grasp the methodological constraints and opportunities of using the Japan Patent Office Database (IIP Patent Database) and compare these results to other widely used databases (the United States Patent and Trademark Office Database - USPTO Database). We aim to bring a more comprehensive discussion on the usability of patent information for geographically oriented analysis and, mainly, raise the attention of other scholars on the challenges in working with raw spatial data. Our discussion identifies constraints regarding the inventor's address data and patent technology classification. Similarly, we propose opportunities for future research that value the possibilities for replicability (free-use databases) and highlight some solutions for the constraints. Additionally, our spatial autocorrelation analysis found only positive autocorrelation in one technology domain. We discuss this result in light of the raw database cleaning method and analysis.

Key words Patent database, Technology, Innovation, Spatial autocorrelation, Japan

Introduction

Literature on innovation in Geography has long focused on analyzing the spatial distribution of innovation activities. Identifying areas in which innovation activity concentrates occupied economic geographers and the innovation studies literature for decades (Audretsch and Feldman 1996; Bouba-Olga 2005). The study of such spatial distributions took several approaches in the literature, from relational and network studies of knowledge diffusion in local and transnational settings (Bathelt et al. 2011; Glückler et al. 2017; Petrov 2011) to cluster creation and evolution (Bathelt et al. 2004; Malmberg and Maskell 2002).

Data for analyzing such research is also varied. Quantitative and qualitative methods proved to help understand how and where innovation occurs and the dynamics of such patterns (Balland 2012; Butzin and Widmaier 2016; Liefner and Jessberger 2016). Among these methods, considerable research used patent information to grasp innovation's spatialization, sectorial analysis, and dynamics (Acs et al. 2002; Crescenzi et al. 2022, 2020; Ejermo 2009; Erdem and Mert Cubukcu 2022; Jaffe et al. 1993; Nomaler and Verspagen 2016; Rigby 2015; Sonn and Storper 2008; Stek 2020).

The following article presents some methodological constraints and opportunities in using patent information to analyze technology concentration patterns in Japan. We focus our analysis on comparing results in processing and measuring geostatistical indicators using the Institute of Intellectual Property Database (hereafter IIP Patent Database[1]) from Japan and the United States Patent and Trademark Office (USPTO) Database. Both databases are suitable for conducting studies, validating their usability in different contexts (Kim and Lee 2015).

We use the Local Moran's I indicator of spatial autocorrelation to identify clustering patterns using both databases and the local indicator of spatial association (also known as LISA) to identify cluster locations (Anselin 1995). Additionally, we use Goto and Motohashi's (2007) classification of technological domains applied to the Japanese database to categorize technologies. The results show few concentration patterns for most technological domains, except Chemicals (using the IIP Patent Database), and low concentration patterns in all categories using USPTO Database.

Patent Information for Innovation Research in Economic Geography

In the innovation and Economic Geography literature, many studies have utilized patent information as a proxy for innovation performance at various scales (Lissoni and Miguelez 2014). Patent data can be utilized to acquire comprehensive information on the loci of innovation, the affiliations of inventors, and the proprietary rights thereof. It can be employed to discern where relevant technologies are clustered, thereby facilitating discussion regarding the strengths and weaknesses of a specific region's economy (Nomaler and Verspagen 2016; Rigby 2015; Stek 2020). Additionally, it can capture the flow of knowledge and technology between regions, which allows us to show the dynamics of innovation networks, not only locally but also globally (Breschi and Lissoni 2009; Erdem and Cubukcu 2022; Jaffe et al. 1993; Sonn and Storper 2008).

However, there are several drawbacks to using patent data in economic geography and other innovation studies. Firstly, patent data is not representative of all innovations and can only capture those coded as patents (Griliches 1979; Pakes and Griliches 1980). In addition, depending on the industry or type of business, there are cases where many patents are applied for, while there are cases where no patents are applied for in order not to make their public information (Moser 2012; Ziedonis 2004). It is challenging to learn about these patenting strategies from databases. Furthermore, patent information, especially addresses, is not optimized for analysis, which often poses difficulties when analyzing complex data (Goto and Motohashi 2007). This issue will be discussed in more depth later. When utilizing patents as a proxy for innovation, the aforementioned considerations must be taken into account.

Compared to US and Europe, Japan has lagged behind in the development of patent databases, and patent data has not been fully utilized in innovation research (Suzuki and Goto 2007). Although there have been studies in economic geography and adjacent fields using Japanese patents (Kamakura 2014; Lim and Kidokoro 2013; Mizuno 2004, 2001), the analysis has been limited to specific time periods or specific firms. More recently, Takeuchi et al. (2018), using the IIP Patent Database 2015, an earlier version of the database used in this paper, found that technological networks are concentrated at close geographic distances, with inventors in different regions cooperating at smaller time distances. Koyanagi (2021) attempts to analyze industry-academia collaboration using the same IIP Patent Database 2020 as this paper, and attempts to address the issue of address data by creating his own dictionary of place names in his analysis.

The IIP Patent Database contains

information on patent applications, applicants, inventors, rights holders, and citations. Goto and Hashimoto (2007) deeply analyzed the characteristics of the IIP Patent Database, founding the main constraints to the Japanese patent data. For instance, the authors explained that applicant information was not systematized until 1992, creating information lacunas in many fields, such as addresses. This is also true for the inventor data, which is the primary information source for spatializing patent data (Ó hUallacháin 2012).

Despite its problems, the IIP Patent Database has been widely used in the spatial and sectorial analysis. For instance, Matsumoto (2017) conducted a panel analysis (fixed effects model) using spatial data at the municipal level with attribute information from 20 years (1993-2012) of Japanese patent data from the IIP Patent Database. The results suggest that in the relationship between Marshall-Arrow-Romer (MAR) externalities and innovation patterns, MAR externalities tend to work better for technologies with lower technology occupations and higher entry barriers. Similarly, in the relationship between Jacobs' externality and innovation patterns, it is suggested that Jacobs' externalities tend to work more readily for technologies with lower technology occupation and higher entry barriers.

The IIP Patent Database is widely recognized as a primary source of patent information for research purposes and is commonly used in conjunction with other commonly utilized databases such as the USPTO, EPO, and KIPO databases. While these databases provide valuable insights into innovation output and performance, it is essential to note that their results may vary due to the specific regions they cover and the fact that patent applications are often filled with the nearest patent office to the applicant. This is highlighted in the literature when comparing the database's spatial scope (Kim and Lee 2015).

Using patent information provides a proven glance into technology and innovation distribution in space. In this article, we use patent information as an exercise to compare the Japanese and the US databases while also introducing geostatistical analysis to identify agglomeration patterns at the technology level. In the following section, we detail the steps and the analysis used to illustrate those differences. We pay particular attention to the database cleaning process, as this step better resumes the constraints of working with raw databases.

Methodology

The main focus of this article is to present the common problems in working with the IIP Patent Database for mapping and geostatistics purposes. In this section, we resume the characteristics of the data, the tidying-up process, and the spatial statistics used to compare the results between the IIP Patent Database and USPTO Database. The database processing uses R 4.2.2, and for mapping, we use ArcMap 10.7. Also, we utilize GeoDa Software for the geostatistics analysis.

The data

The IIP Patent Database contains five files connected by a key field (*ida*), the patent application number. As presented in the user manual, the database contains around 14.3 million patent applications between 1964 and 2019 (see Table 1). The IIP Patent Database

Table 1. IIP Patent Database resume.
IIP Patent Database User Manual.

Table Name	File Name	Number of Records
Application	ap.txt	14,303,616
Inventor	inventor.txt	29,601,494

Source: Own elaboration.

is one of the patent databases with the longest records, this helps analyze significant trends in patent applications and technologies over time.

This analysis uses the application file to identify technology classes following the International Patent Classification (IPC), patent application date, and registered patents. To map the patents, we use the inventor's address as proposed by other studies that georeferenced patent information (Ó hUallacháin 2012; Stek 2020). Inventor information is preferred over applicant information because it grasps better where the inventions take place, as applicant and IP holder addresses usually refers to the company or institution where the inventor works. However, as the inventor address field is not standardized, it is common to find company/institution addresses in this field. This problem is unavoidable in this database, as the registered information is inconsistent across cases. Other studies (Ó hUallacháin 2012) argue that workers usually live close to their workplace, so the distortion caused by registering institution/company addresses as inventor addresses is not particularly relevant. Nevertheless, this problem could bring some concerns for the Japanese case, as city and municipality boundaries are much smaller than in the United States, particularly in highly populated cities. This might cause an unusual concentration of patent activities in areas with low populations and high company/institution headquarters presence (Mizuno 2022).

The selection of the municipality scale is not arbitrary, as the prefecture level might not capture subtle territorial differences between cities. However, these differences can be captured within the city limits in big urban agglomerations like Tokyo, Osaka, or Nagoya. Other studies mentioned above, like Stek (2020), use georeferenced information at local inventor addresses to create a global innovation heatmap. Aggregating information at the municipality level is helpful because many addresses in the database are incomplete, presenting only partial address information.

The USPTO Database is differently organized in this regard. There are several sources of information in the USPTO Database. We used the patents view platform, which allows bulk database download for registered patents (not patent applications[2]). The data used for the analysis include the "g_patent" file, which includes the general patent date, the key ID (*patent_id*), and the "g_location_disambiguated" (that holds location information), "g_cpc_title" to identify the patent's technology domain and the "g_inventor_disambiguated" that contains the inventor's information (see Table 2). The database contains patents from 1976 to 2022.

The patent data utilized in this analysis is free to access. Therefore, results are reproducible by other researchers interested in patent data mapping and analysis. However, first, the raw information in this database must be cleaned and prepared before being used for the analysis. This process was particularly relevant for the Japanese case as much information was incomplete or wrong, mainly when using the address as a source of locational information.

Tidying up

Raw databases present various challenges before utilizing them. Both databases contain more or less organized raw data, as presented

Table 2. USPTO Database resume.

Table Name	File Name	Number of Records
Patent	g_patent.tsv	8,169,776
Location	g_location_disambiguated.tsv	32,715,446
CPC Title	g_cpc_title.tsv	264,485
Inventor	g_inventor_disambiguated.tsv	20,159,421

Source: Own elaboration.

in the previous section. The cleaning process contains two main steps for both databases. First, select the inventor cases that contain address information in Japan. Second, combine the selected cases with technology classification information from the patent file (g_patent). With those two sources, we can relate addresses to municipalities (for mapping purposes) and classify technological categories for further representation.

Cleaning and extracting address information is possibly the most complicated step in working with the Japanese database. As previously mentioned, address information is not standardized and presents particular problems when extracting prefecture and municipality. For instance, many records have no prefecture or municipality information, while others are not precise (mistyping). Therefore, we first tried extracting municipality names and prefectures using the Zipangu Library in R. The library contains a list of topographic and administrative boundary names used in conjunction with a piece of code that extract prefecture names, city/municipality name, and the rest of the address.

The code creates a function that separates Japanese addresses into three different columns. The problem with using the Zipangu library is that the names list is restricted to the format Prefecture-City-rest of the address, and therefore, the matching strategy (searching within the string) does not necessarily return city or prefecture names even when these are in the string[3]. Also, the names list in the Zipangu Library only contains recent city and prefecture names. As Japanese cities have been merged, renamed, and transformed over the last decades (Suzuki and Sakuwa 2016), many names are not captured by the algorithm or are wrongly retrieved.

To overcome this first difficulty, we utilized a list of cities available from 1970 to 2019. The Municipality Map Maker (MMM), prepared by Takashi Kirimura[4], summarizes municipality changes over time. This list makes it easy to track municipality mergers and other transformations. Using the city and prefecture names list, we search within each case for string matches between the database and the list. However, this uses substantial computational power with processing times of up to 28 hours for the 29 million registries. However, to reduce computing time and comparability with the USPTO Database, we filtered registered patents for approximately 9.2 million cases.

Once we obtained prefecture and city names, we merged patent information with inventor information to identify the patent location. As a patent usually has multiple inventors, the location of a patent is determined by the proportion of each inventor[5]. Unlike other similar analyses that assign the location only to the first inventor (Ó hUallacháin 2012), we believe that in the Japanese case in very populous cities, inventors might not live within the same municipality. Therefore, it is essential to differentiate the patent location proportionally from the number of inventors.

For the USPTO Database, the cleaning process is much simple. Location information is given in coordinates. Additionally, a field in the inventor table identifies the inventor's nationality. Filtering Japanese inventors is easy; however, location coordinates do not always refer to Japanese coordinates, as many points do not reflect the municipality's name. To solve this problem, we manually checked the mapped coordinates (points) with their polygon counterpart (municipalities shapefile). A simple join between the two data confirms the correct matches.

Technology information is similar too. For example, the CPC Title table contains CPC codes for each patent. By joining data tables from location and technology, it is easier to construct a shorter database that contains the necessary information for mapping patents and classifying technologies. To classify technologies, we used Goto and Motohashi's (2007) technology classification, which is based on the WIPO NBER (National Bureau of Economic Research) (see Table 3).

The next step is to map and identify technology clusters in Japan. Using the NBER classification as a macro-category, we run geostatistical tests to identify spatial autocorrelation and clustering. In the next section, we briefly detail the statistics used.

Geostatistics

As part of the exploratory exercise to identify clusters and outlier values, we run a local indicator of spatial association (LISA), which includes the Local Moran's I indicator (Anselin 1995). However, before computing LISA, we standardized the number of patents in a technological category by the municipality size, measured by habitants. This is necessary due to the agglomeration tendency in areas with higher population rates (Ó hUallacháin 2012).

The Local Moran's I indicator of spatial autocorrelation is one of the most popular

Table 3. IPC Classification

NBER Name	Title	IPC Codes
Chemical	Separating, mixing	B01, B02, B03, B04, B05, B06, B07, B08, B09
Chemical	Non-organic chemistry, fertilizer	C01, C02, C03, C04, C05
Chemical	Organic chemistry, pesticides	C07, A01N
Chemical	Organic molecule compounds	C08
Chemical	Dyes, petroleum	C09, C10, C11
Computer and Communications	Clock, comtrolling, computer	G04, G05, G06, G07, G08
Computer and Communications	Display, information storage, instruments	G09, G10, G11, G12
Computer and Communications	Electronics circuit, communication tech.	H03, H04
Pharmaceuticals and Medical	Health and amusement	A61, A62, A63
Pharmaceuticals and Medical	Drugs	A61K
Pharmaceuticals and Medical	Biotechnology, beer, fermentation	C12, C13, C14
Pharmaceuticals and Medical	Genetic engineering	C12N
Electrical and Electronics	Measurement, optics, photography	G01, G02, G03
Electrical and Electronics	Nuclear physics	G21
Electrical and Electronics	Electronic components, semiconductor	H01, H02, H05
Mechanical	Machine tools, metalworking	B21, B22, B23
Mechanical	Casting, grinding, layered product	B24, B25, B26, B27, B28, B29, B30, B32
Mechanical	Transporting	B60, B61, B62, B63, B64
Mechanical	Packing, lifting	B65, B66, B67, B68
Mechanical	Metallurgy, coating metals	C21, C22, C23, C24, C25, C26, C27, C28, C29, C30
Mechanical	Engine, pump	F01, F02, F03, F04, F15
Mechanical	Engineering elements	F16, F17
Other	Agriculture	A01
Other	Food Stuff	A21, A22, A23, A24
Other	Personal and domestic articles	A41, A42, A43, A44, A45, A46, A47
Other	Printing	B41, B42, B43, B44
Other	Textile	D01, D02, D03, D04, D05, D06, D07
Other	Paper	D21, B31
Other	Construction	E01, E02, E03, E04, E05, E06, E21
Other	Mining, drilling	E21
Other	Lighting, steam generation, heating	F21, F22, F23, F24, F25, F26, F27, F28
Other	Weapons, blasting	F41, F42, C06
Other	Others	B82

Source: Based on Goto and Motohashi (2007).

indexes to measure the concept of spatial autocorrelation. This concept is particularly relevant in Geography and rarely discussed in depth, but in simple terms, it is a measure of correlation within georeferenced variables across space (Getis 2008). The spatial autocorrelation concept uses a geographic-based principle; spatial phenomena are not independent, and their relation decays with distance. The Local Moran's I indicator uses the global examination of Moran's I indicator first proposed by Moran (1950, 1948). The index reveals if an observation has similar characteristics to its neighbors for a univariate variable. The spatial autocorrelation can be positive or negative. Positive when there is a tendency to be similar (can indicate clustering) or negative when there is no evidence of neighbors with similar characteristics. The result is a scatterplot divided into four regions; observations with high-high correlation (high values surrounded by high values), low-low (low values surrounded by low values), high-low (high values next to low values), and low-high (low values next to high values). Within these combinations, high-high and low-low are cases in which spatial autocorrelation is positive. This, however, only indicates the spatial autocorrelation degree but does not reveal clustering information. Therefore, the analysis is complemented with LISA maps which show cluster locations and their significance (see Figure 1 and 2).

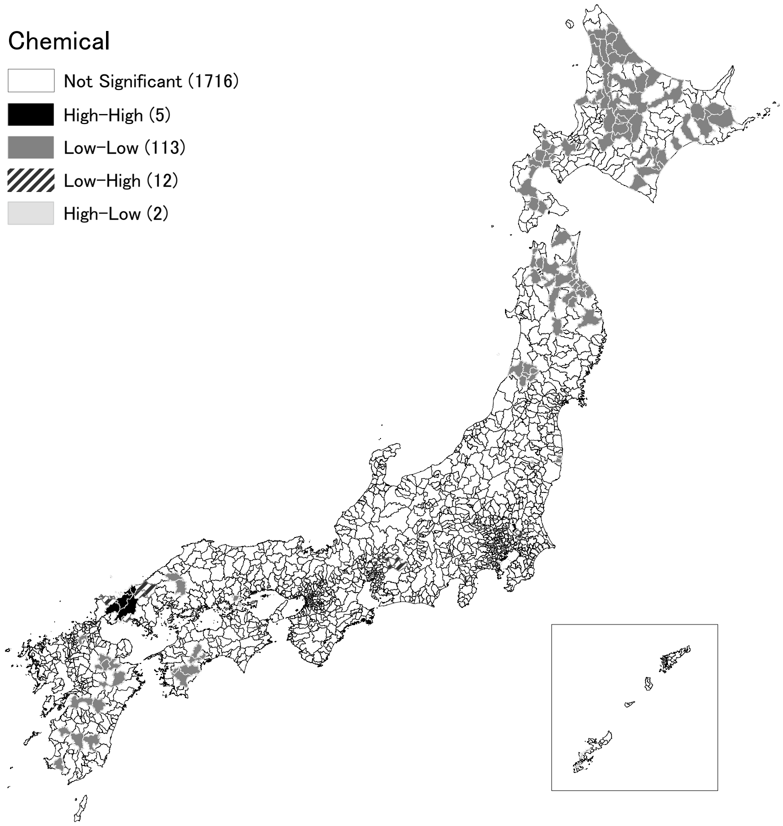

Figure 1. LISA Maps of spatial autocorrelation for Chemicals using the IIP Patent Database.
Source: Own elaboration.

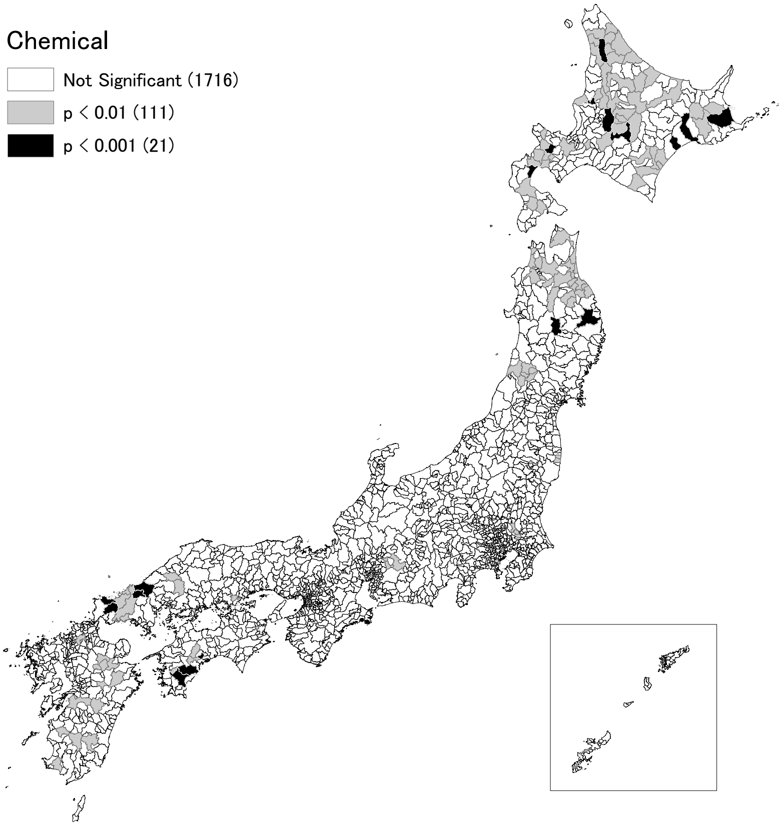

Figure 2. Significance coefficient for Chemicals using the IIP Patent Database.
Source: Own elaboration.

A relevant aspect of spatial autocorrelation is carefully selecting the weight matrix for neighbor selections. There are multiple ways to do this, whether defining a euclidean distance from which any other observations will count as neighbors by stating the number of neighbors for every observation (K-neighbors). In our exercise, we use the queen contiguity measure to define neighbors, which is the most used when shapes are irregular, and it is the default option for constructing weight matrices in GeoDa. The Queen criterion defines neighbor polygons that share vertices and sides with the main observation.

Our weight matrix has a maximum of 15 neighbors for observation, a mean of 5.14, and 48 neighborless observations (see Figure 3). Neighborless observations are relevant, as Moran's I index does not consider them in the calculation. Therefore, the final computation includes both databases, using the same weight matrices and shapefile. A results review is presented in the following section.

Results

After running the tests for clustering and spatial autocorrelation, we found, in general, a lack of evidence to suggest that technological inventions clusters within the Japanese context. Furthermore, Moran's I indicator of spatial autocorrelation is low in all categories

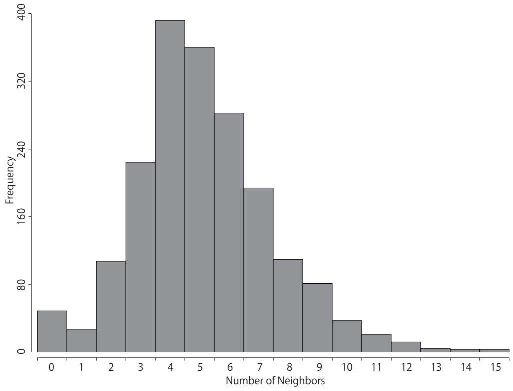

Figure 3. Queen contiguity histogram.
Source: Own elaboration.

except for chemicals using the Japanese database (see Figure 4), when the indicator is higher. Even eliminating outlier data does not raise the spatial autocorrelation indicator, evidencing a calculation problem or no spatial autocorrelation. Utilizing the USPTO Database does not show any spatial autocorrelation either (see Figure 5). While data presents more outliers, eliminating them slightly increases the spatial autocorrelation index in some technology domains.

It is relevant to note that using the IIP Patent Database includes more municipalities without null values, as there are more patent data for each municipality. On the contrary, the USPTO Database primarily includes patents registered with an address in Tokyo (e.g., Chiyoda City in Tokyo - data outliers), bringing back the problem of company address registration as an inventor address. This is a known problem that is more evident in foreign patent registration.

However, it is surprising that the spatial autocorrelation index is much lower than in other examples that use patent data for technological clusters (Ó hUallacháin 2012). While this is not the only methodology used for proximity analysis and technology specialization (Khramova et al. 2013), this widely used method permits comparisons between previous studies within economic geography.

As Chemicals using the IIP Patent Database is the only result with a slightly higher spatial autocorrelation index, we can use this result to map and identify clusters using the LISA method. For example, Figure 1 shows that at 0.01 significance level, the areas with the high significance of High-High spatial autocorrelation are located in Yamaguchi Prefecture. In particular, Yamaguchi-shi, Hagi-shi, Mine-shi, and Abu-cho. This area is known for Chemicals and other manufacturing industries. In the Census of Manufactures 2020, Yamaguchi Prefecture ranks first in Japan in the value of manufactured goods per plant, particularly in the chemical industry. Another area that shows High-High spatial autocorrelation is Joso-shi in Ibaraki Prefecture.

On the other hand, places with Low-Low spatial autocorrelation are found mainly in Hokkaido, Aomori, Iwate, and Akita Prefectures in the North. Some municipalities in Kochi, Oita, Miyazaki, and Kagoshima Prefectures are in the South.

Discussion

Constraints

Using patent information as a proxy of innovation performance and technological output is still problematic. Some studies have discussed whether patent data helps indicate innovation success (Reeb and Zhao 2020), while others have questioned using patent data for measuring technological trajectories (Filippin 2021). Despite such legitimate discussions over the limits of patent data, the methods that use it as a primary source of information must consider its constraints and problematics.

In our exercise, we encounter numerous problems using patent data from raw databases. We can identify two main issues: incomplete

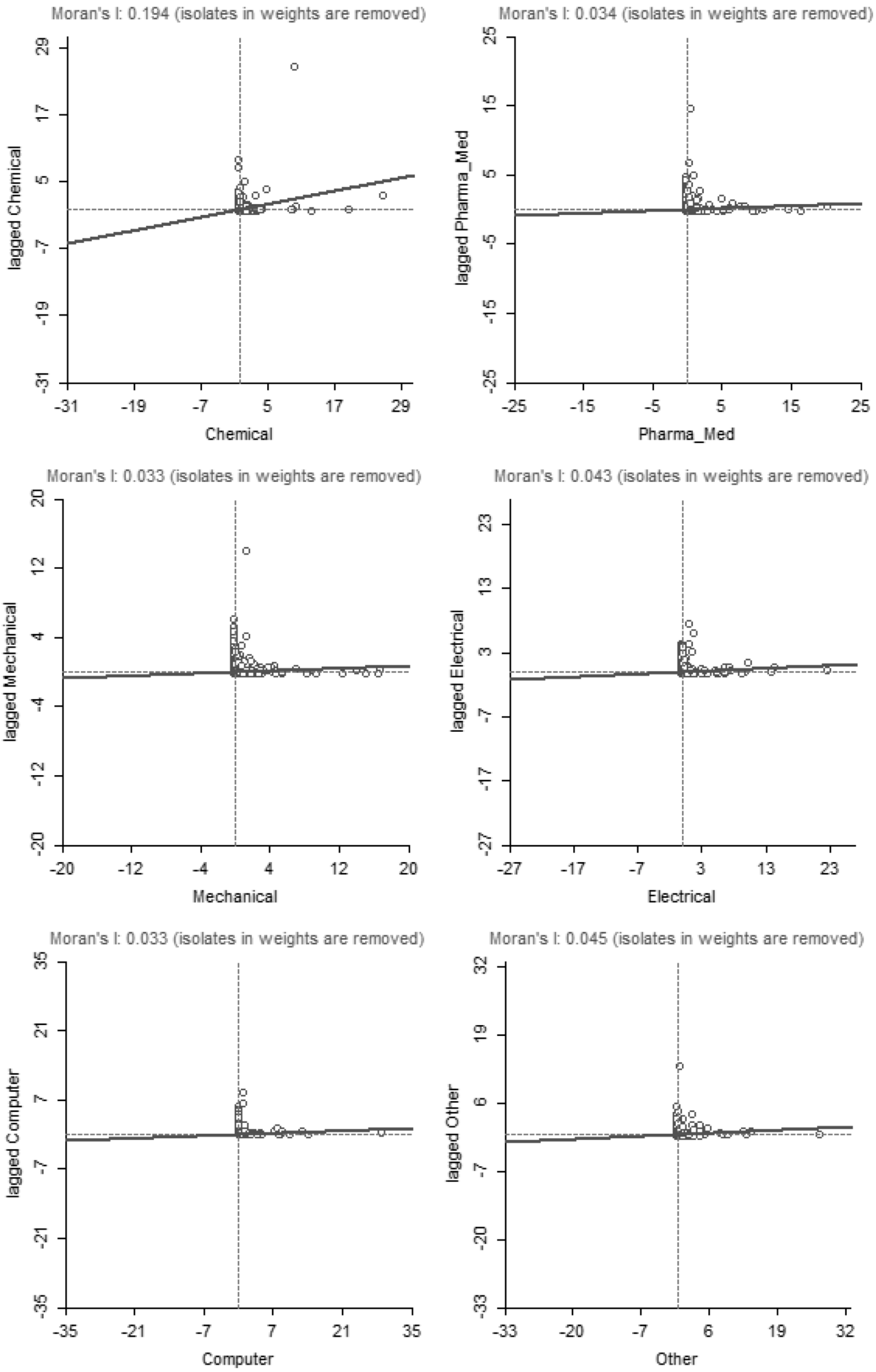

Figure 4. Moran's I Scatter Plot using IIP Patent Database.
Source: Own elaboration.

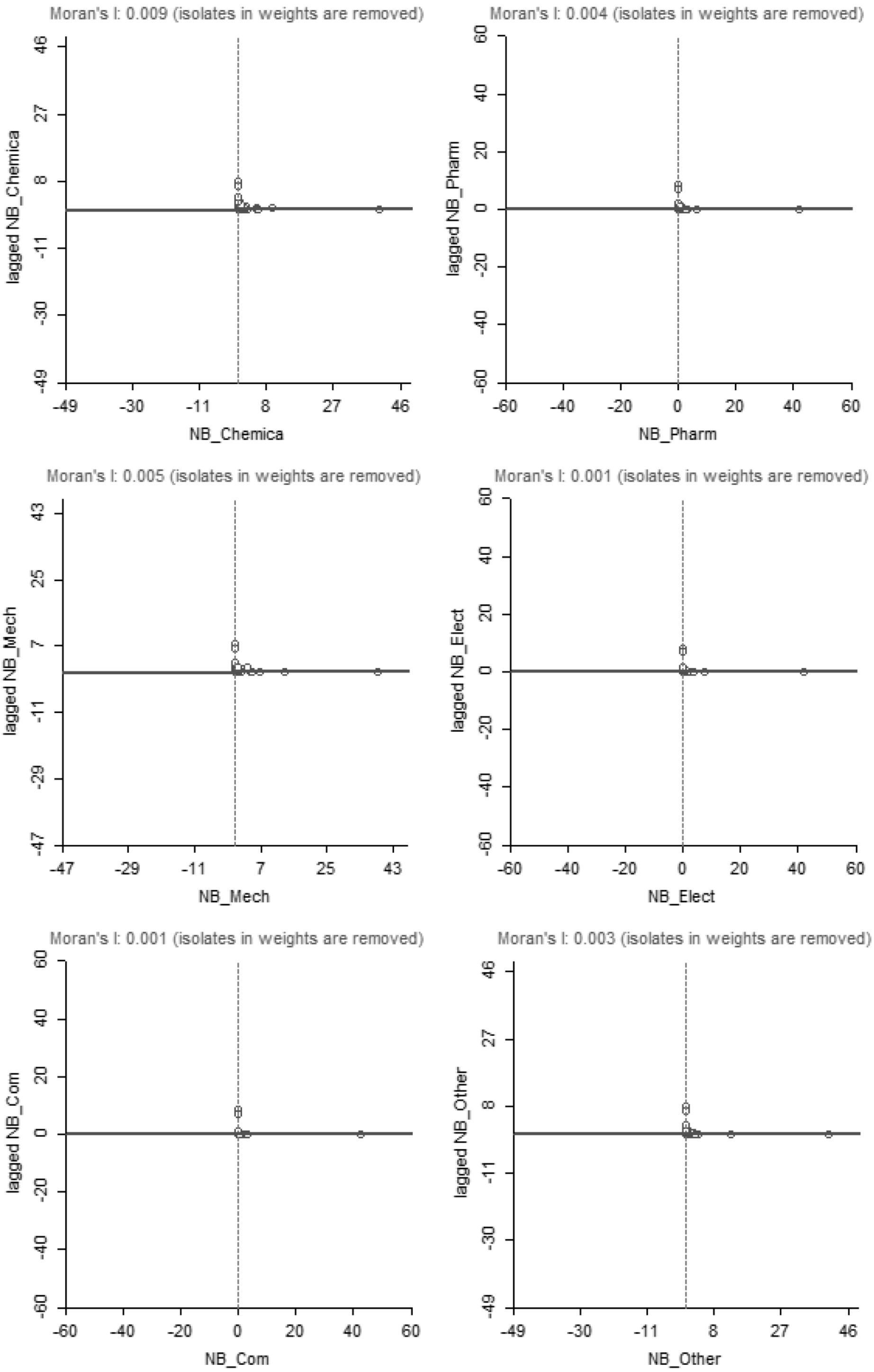

Figure 5. Moran's I Scatter Plot using USPTO Database.
Source: Own elaboration.

or wrong spatial data and patent database relevance. First, during the methodology section, we identified problems in working with address data, particularly in the IIP Patent Database case. The incomplete or non-existing address data or the fact that many inventors use company information when registering patents makes it challenging to clarify the location where inventions occur. This impacts not only this short analysis but is also an inevitable and inherent part of the database.

Additionally, extracting address information for mapping can carry various mistakes. For instance, many municipality names are similar to others in other prefectures or toponyms. Therefore, extracting municipality and prefecture names from the address field can involuntarily extract the wrong data, leading to misinterpretation. Other studies use geocoding to solve this problem. For instance, Stek (2020) utilizes the Google Maps API to map inventors' addresses into point data that can be transformed into heat maps. While this technique can also lead to mapping errors, geocoding is commonly used when working with geographic data.

Secondly, there is a persistent view that most patents do not capture innovative or technological advancements well (Dang and Motohashi 2015; Ejermo 2009; Reeb and Zhao 2020). Some of these studies argue that patent quality might be a problem and work on measures to select patents that can truly evidentiate technological leaps (Squicciarini et al. 2013). In our exercise, we selected only granted patents for both databases to overcome this problem. However, this step does not ensure that selected patents best represent the inventiveness of a region or municipality. Related to this problem is the selection of technology categories for the sectorial analysis; as many patents have two or more classifications, it is not easy to classify patents into a single category. We partially solve this by selecting the first code inscribed in the registry. However, this does not ensure that a single patent can be applied to multiple technological macro-categories. Related to this issue is that technology classification categories might exclude some municipalities, making their value zero. However, this does not mean that these municipalities do not have patents; it signifies that some municipalities do not have patents in the selected categories. This might eventually affect the spatial autocorrelation index result and consistency.

These constraints make it challenging to capture the accurate spatial dimension of technology development. While most are methodological decisions, this can make a substantial difference when calculating spatial statistics or using other indicators to identify clusters.

Opportunities

Apart from identifying particular constraints in working with patent data, we also want to highlight the possibilities and ways to improve the analysis for future iterations. The proposed improvements align with the constraints identified but also highlight the relevance of patent information as an introductory source of information for further research.

First, we would like to emphasize that despite its evident challenges, mapping addresses provide major opportunities for more flexible analysis. For instance, addresses allow rapid-scale changes in the data as the information is not inherently tied to a particular scale. The possibilities for zooming in and out within prefectures or other macro-regions represent an opportunity for sub-national level analysis, an issue that concerns present patent research (Crescenzi et al. 2020). Such flexibility implies improving how we extract valuable data for mapping purposes. Here, our exercise proves that occupying a list of common names can bring relevant results, but it is more accurate to geocode addresses. The problem with geocoding techniques is that they require high computational power or are prohibitively expensive. As patent databases contain millions of entries, geocoding becomes a resource challenge.

Second, we propose using Principal Component Analysis (PCA) to ensure a more "organic" technological classification adapted to the national context. Other studies in this field have used PCA to reduce technological categories (Ó hUallacháin 2012). This implies that one can identify sectorial groups that evidence contextual differences depending on the data used. Constructing a sectorial classification that grasps local particularities is highly relevant to the Japanese case, as the demographic and other geographical characteristics (e.g., higher population densities or urban agglomerations configurations) might differ from other occidental contexts.

Moreover, we suggest using patent data as an exploratory and introductory way to other methodologies for analyzing local technological inventiveness or innovation performance. For instance, combining indicators from the viewpoint of organizational proximity (Boschma 2005; Kaygalak and Reid 2016; Sonn and Storper 2008), for example, might produce positive synergies to evaluate multiple proxies into innovation and regional economic performance.

Finally, we believe that raw and freely available data helps reproducibility (Baker 2016). Therefore, it is necessary to open the discussion into which data we are using as social scientists and the possibilities for future research to reproduce experimental results. With quantitative information, this is much easier to accomplish. Despite its difficulties and problems, using raw patent data from free access resources is an excellent step in this direction; It poses a more sincere discussion on data usability, sources, and methods.

Conclusions

Studies on innovation and technology using patent information are common. A wide range of research utilizes patent information for analyzing technological paths, technology clusters, and regional innovation performance. From economic geography to innovation studies and economics, patent databases have been proven to help answer various questions and topics. In this brief exercise, we evaluate the constraints and opportunities that patent databases offer to map and analyze technology clusters in Japan.

We found that working with address data in the case of the IIP Patent Database might cause some problems in extracting and validating the information. The lack of standardized address information as an input cause trouble in using it as a source for geographical reference. In the case of the USPTO, the information is more organized; however, it lacks capturing diverse locations as they mainly refer to company headquarters rather than the inventor's home address.

We also open the question of whether patent information helps capture sectorial differences, as they might not grasp relevant technological advancements or other technological innovations. Again, this discussion is not particularly new, but it is crucial to revisit such claims to comprehend the limitations of using patent data and to inform methodological decisions better.

We suggest that some improvements must be made to illustrate accurate technological clustering. First, enhance information extraction or use geocoding for mapping address data; second, we propose using PCA analysis for technology classification to capture the particularities of the Japanese case in future research. Finally, further steps must be taken toward a more comprehensive quantitative analysis of technology regions. This implies exploring other methodologies and indicators for in-depth analysis of the multiple dimensions of innovation and regional development.

Acknowledgements

We are grateful to the Institute of Intellectual Property for providing us with the IIP Patent Database. This research was supported by JSPS KAKENHI Grant Number 20K13266.

Notes

1. The IIP Patent Database is available upon request in the following link: https://www.iip.or.jp/e/patentdb/index.html . After revision of the submitted formulary, you will receive the download links and the aforementioned user manual.
2. The tables used from the Patent View Platform can be found in the following link: https://patentsview.org/download/data-download-tables
3. The function was posted in https://exploratory.io/note/1021500949444839/vrF1QEF5hR/note_content/note.html . Credits to the author.
4. Kirimura's database contains also shapefiles useful for mapping municipalities across time. Please visit it in http://www.tkirimura.com/mmm/
5. For instance, if a patent X has three inventors with address in three different municipalities, each municipality gets 1/3 of the patent.

References

Acs, Z. J., Anselin, L., and Varga, A. 2002. Patents and innovation counts as measures of regional production of new knowledge. *Research Policy* 31: 1069–1085.

Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27: 93–115.

Audretsch, D. B. and Feldman, M. P. 1996. R&D spillovers and the geography of innovation and production. *The American Economic Review* 86: 630–640.

Baker, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452–454.

Balland, P.-A. 2012. Proximity and the evolution of collaboration networks: Evidence from research and development projects within the Global Navigation Satellite System (GNSS) Industry. *Regional Studies* 46: 741–756.

Bathelt, H., Feldman, M. P. and Kogler, D. F. eds. 2011. *Beyond territory: Dynamic geographies of knowledge creation, diffusion, and innovation*. New York: Routledge.

Bathelt, H., Malmberg, A. and Maskell, P. 2004. Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28: 31–56.

Boschma, R. 2005. Proximity and innovation: A critical assessment. *Regional Studies* 39: 61–74.

Bouba-Olga, O. 2005. Innovation, proximities and regional development. *European Journal of Economic and Social Systems* 18: 285–305.

Breschi, S. and Lissoni, F. 2009. Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography* 9: 439–468.

Butzin, A. and Widmaier, B. 2016. Exploring territorial knowledge dynamics through innovation biographies. *Regional Studies* 50: 220–232.

Crescenzi, R., Dyèvre, A. and Neffke, F. 2022. Innovation catalysts: How multinationals reshape the global geography of innovation. *Economic Geography* 98: 199–227.

Crescenzi, R., Iammarino, S., Ioramashvili, C., Rodríguez-Pose, A. and Storper, M. 2020. The geography of innovation and development:

Global spread and local hotspots. Geography and Environment Discussion Paper Series 4.
Dang, J. and Motohashi, K. 2015. Patent statistics: A good indicator for innovation in China? Patent subsidy program impacts on patent quality. *China Economic Review* 35: 137–155.
Ejermo, O. 2009. Regional innovation measured by patent data—Does quality matter? *Industry and Innovation* 16: 141–165.
Erdem, U. and Cubukcu, K. M. 2022. The uneven geography of innovation in Turkey: Visualizing the geography and regional relatedness of patent production. *Environment and Planning A* 54: 7–10.
Filippin, F. 2021. Do main paths reflect technological trajectories? Applying main path analysis to the semiconductor manufacturing industry. *Scientometrics* 126: 6443–6477.
Getis, A. 2008. A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographical Analysis* 40: 297–309.
Glückler, J., Lazega, E. and Hammer, I. eds. 2017. *Knowledge and networks*. Cham: Springer International Publishing.
Goto, A. and Motohashi, K. 2007. Construction of a Japanese patent database and a first look at Japanese patenting activities. *Research Policy* 36: 1431–1442.
Griliches, Z. 1979. Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics* 10: 92–116.
Jaffe, A. B., Trajtenberg, M. and Henderson, R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108: 577–598.
Kamakura, N. 2014. Organizational restructuring and locational hysteresis in R&D: Case study of spatial division of labor in Japanese synthetic chemical companies established by *zaibatsu*. *Geographical Review of Japan series A* 87: 291–313. (JE)
Kaygalak, I. and Reid, N. 2016. Innovation and knowledge spillovers in Turkey: The role of geographic and organizational proximity. *Regional Science Policy & Practice* 8: 45–60.
Khramova, E., Meissner, D. and Sagieva, G. 2013. Statistical patent analysis indicators as a means of determining country technological specialisation. Higher School of Economics Research Paper No. WP BRP 09/STI/2013.
Kim, J. and Lee, S. 2015. Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. *Technological Forecasting and Social Change* 92: 332–345.
Koyanagi, S. 2021. Tokkyo kyodo shutsugan ni miru kenkyu kaihatsu no kukan bunseki: Sangaku renkei wo chushin ni. Presentation at Southwest branch meeting of the Japan Association of Economic Geographers. (J)
Liefner, I. and Jessberger, S. 2016. The use of the analytical hierarchy process as a method of comparing innovation across regions: The examples of the equipment manufacturing industries of Shanghai and Xiamen, China. *Environment Planning A* 48: 1188–1208.
Lim, H. and Kidokoro, T. 2013. The spatial characteristic of innovation network through analysis of joint patent network in Japan. *Journal of the City Planning Institute of Japan* 48: 567–572. (JE)
Lissoni, F. and Miguelez, E. 2014. Patents, innovation and economic geography. *The WIPO Journal* 6: 37–49.
Malmberg, A. and Maskell, P. 2002. The elusive concept of localization economies: Towards a knowledge-based theory of spatial clustering. *Environment Planning A* 34: 429–449.
Matsumoto, K. 2017. *Tokkyo shutsugan wo chushin toshita inobeshon niokeru shuseki no gaibusei ni kansuru jissho kenkyu*. Doctoral dissertation, The University of Tokyo. (J)
Mizuno, M. 2001. Technological innovations, inter-firm networks and distance: A geographical analysis of patent data in the Japanese

automobile industry. *Japanese Journal of Human Geography* 53: 18–35. (JE)

Mizuno, M. 2004. Technological innovations and geographic proximity in interfirm networks: A case study of small and medium-sized enterprises in Osaka Prefecture. *Geographical Review of Japan*, 77: 940–953. (JE)

Mizuno, M. 2022. Why do firms concentrate in Tokyo? An economic geography perspective. *Japan labor Issues* 6(37): 43–54.

Moran, P. A. P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)* 10: 243–251.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23.

Moser, P. 2012. Innovation without patents: Evidence from world's fairs. *The Journal of Law & Economics* 55: 43–74.

Nomaler, Ö. and Verspagen, B. 2016. River deep, mountain high: Of long run knowledge trajectories within and between innovation clusters. *Journal of Economic Geography* 16: 1259–1278.

Ó hUallacháin, B. 2012. Inventive megaregions of the United States: Technological composition and location. *Economic Geography* 88: 165–195.

Pakes, A. and Griliches, Z. 1980. Patents and R&D at the firm level: A first report. *Economics Letters* 5: 377–381.

Petrov, A. 2011. Beyond spillovers: Interrogating innovation and creativity in the peripheries. In *Beyond territory: Dynamic geographies of knowledge creation, diffusion, and innovation*. ed. H. Bathelt, M.P. Feldman and D. F. Kogler, 168-190. New York: Routledge.

Reeb, D. M. and Zhao, W. 2020. Patents do not measure innovation success. *Critical Finance Review* 9: 157–199.

Rigby, D. L. 2015. Technological relatedness and knowledge space: Entry and exit of US Cities from patent classes. *Regional Studies* 49: 1922–1937.

Sonn, J. W. and Storper, M. 2008. The Increasing importance of geographical proximity in knowledge production: An analysis of US patent citations, 1975–1997. *Environment Planning A* 40: 1020–1039.

Squicciarini, M., Dernis, H. and Criscuolo, C. 2013. Measuring patent quality: Indicators of technological and economic value, OECD Science, Technology and Industry Working Papers No.2013/03. OECD.

Stek, P. E. 2020. Mapping high R&D city-regions worldwide: A patent heat map approach. *Quality & Quantity* 54: 279–296.

Suzuki, J. and Goto, A. 2007. Innovation research using JPO patent data. *Journal of Intellectual Property Association of Japan* 3: 17–30. (JE)

Suzuki, K. and Sakuwa, K. 2016. Impact of municipal mergers on local population growth: An assessment of the merger of Japanese municipalities. *Asia Pacific Journal of Public Administration* 38: 223–238.

Takeuchi, K., Taima, M., Kidokoro, T. and Seta, F. 2018. Analysis on spatial pattern of Japanese technological innovation based on patent data: Focusing on sectoral innovation system. *Journal of the City Planning Institute of Japan* 53: 172–178. (JE)

Ziedonis, R. H. 2004. Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Science* 50: 804–820.

(J) written in Japanese

(JE) written in Japanese with English abstract