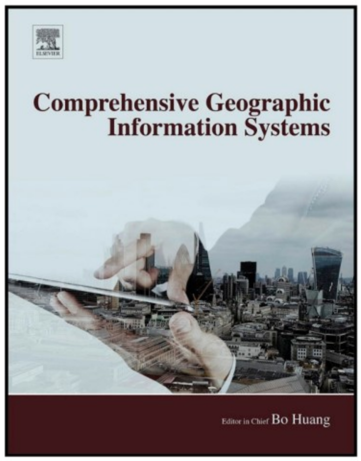

From Ye, X., 2018. Open Data and Open Source GIS. In: Huang, B. (Ed.), Comprehensive Geographic Information Systems. Vol. 1, pp. 42–49. Oxford: Elsevier.
http://dx.doi.org/10.1016/B978-0-12-409548-9.09592-0
ISBN: 9780128046609

1.05 Open Data and Open Source GIS

Xinyue Ye, Kent State University, Kent, OH, United States

1.05.1 Introduction

With the growing capability of recording individual's digital footprints and the emerging open culture, the open big data are flooding everywhere (Batty, 2012). Geospatial data are an important component of open data unfolding right in front of our eyes (Warf and Arias, 2008). Geographic information system (GIS) research is shifting toward analyzing ever-increasing amounts of large-scale, diverse data in an interdisciplinary, collaborative, and timely manner. Goodchild (2013) proposed the crowd, social, and geographic approaches to assess big data quality. Sui (2014) argued that open GIS should involve eight dimensions related to data, software, hardware, standards, research, publication, funding, and education facilitated by web-based tools and the growing influence of the open culture. The key pillars of open GIS have always been and will continue to be open source, open data, open modeling, open collaboration, and open publishing for future GIS research and applications (Rey, 2014). Sui (2014) noted that "the big data torrent will eventually be more powerful if they can be made to conform to open standards such as those developed by OGC over the years". This article places greater emphasis on open data and open source GIS. According to Sui (2014), open GIS offers four exciting opportunities for participation and collaboration among both GIS experts and volunteers: (1) technology-driven opportunities for addressing big data challenges; (2) application-led opportunities for improving decisions across all levels; (3) curiosity-inspired, crowd-powered opportunities for developing citizen science; and (4) education-focused opportunities for realizing a spatial university.

To realize science's powerful capacity for self-correction, it is critical to reproduce the outcomes of scientific research (Ye et al., 2015). However, if data and codes are not transparent, the reproductivity will not work due to bottlenecks or restrictions from copyright, patents, or other mechanisms of control. Open data and open source GIS aim to make GIS research open to everyone. In other words, data and codes should be made legally open and accessible to both professional and nonprofessional communities (Stodden, 2009).

The growing and evident interdisciplinary efforts dedicated to "open data and open source GIS" represent a transformative trend shaped by increased scholarly collaboration and research methods sharing (Sui, 2014). Instead of following the traditional proprietary approach, open data and open source GIS can lead to a large number of benefits to both citizens and businesses across the globe, with the success of GIS research, education, and applications.

The rest of this article is organized as follows. Elements of the emerging open data are described in section "**Open Data**". Section "**Open source GIS**" discusses how open source GIS plays a pivotal role in the research and education. Section "**Practicing Open Source GIS**" demonstrates the use of open source GIS for regional economic analysis. The article ends with a summary and conclusion for open GIS paradigm in section "**Summary**" toward the goal of reproducibility and the desired reuse.

1.05.2 Open Data

Knowledge is eventually derived from data. The term "open data" was coined in 1995 to deal with the disclosure of geophysical and environmental data realizing the idea of common good applied to knowledge (Chignard, 2013). Open data is gaining popularity with the launch of open government data initiatives. The volume of open space–time data in various disciplines and domains has dramatically increased due to the growing sophistication and ubiquity of information and communication technology (Jiang, 2011). Open data can be used and reused at no cost or restriction without mechanisms of control such as copyright and patents (Auer et al., 2007). The intention of open data movement is to make publicly acquired data available for direct manipulation such as cross tabulation, visualization, and mapping (Gurstein, 2011). It is clear that a space–time perspective in using such data has become increasingly relevant to our understanding of socioeconomic and environmental dynamics in the collaborative and transdisciplinary manner. As noted by (Rey, 2014), "Open data constitutes available, intelligible, accessible, and usable data. For science's error–correction mechanisms to kick in, data underlying research projects must be made accessible to the wider research community". Featured by the ever-growing volume, variety, and velocity of ubiquitous geospatial information, the big spatial data in the changing environmental, urban, and regional contexts demand innovative thinking that can capture the rich

Comprehensive Geographic Information Systems, First Edition, 2018, 42–49

information of patterns and processes and provide spatial strategies for sustainable development. Meanwhile, the volume of data created by an ever-increasing number of geospatial sensor platforms such as remote sensing and social sensing (including citizen sensors) to collect data at ever-increasing spatial, spectral, temporal, and radiometric resolutions currently exceeds petabytes of data per year and is only expected to increase. Data come from various sources, types, organizations (governments, military, NGOs, etc.), and purposes. Recent developments in information technology commonly referred to as big data along with the related fields of data science and analytics are needed to process, examine, and realize the value of the overwhelming amount of open geospatial data. The research agenda has been substantially redefined in light of open data, which have transformed the focus of GIS research toward dynamic, spatial, and temporal interdependence of human–environment issues across multiple scales. Metadata is information about data. The use of metadata enhances the opportunities for semantical interoperability involving open data, lowering the cost of access, and manipulating and sharing across data boundaries (Nogueras-Iso et al., 2004). As declared by (FGDC 2017), "Geospatial metadata describes maps, Geographic Information Systems (GIS) files, imagery, and other location-based data resources. The FGDC is tasked by Executive Order 12906 to enable access (see GeoPlatform.gov) to National Spatial Data Infrastructure (NSDI) resources and by OMB Circular A-16 and the A-16 Supplemental Guidance to support the creation, management, and maintenance of the metadata required to fuel data discovery and access".

Twelve critical factors were identified to city-level, regional, and transnational cases regarding the publication and use of open data (Susha et al., 2015): (1) a national guide on legal intellectual property right issues; (2) a clear data publishing process; (3) addressing of societal issues and publishing of related data; (4) interest for users; (5) where to publish datasets; (6) a virtual competence center for technical help; (7) a strategy for maintaining published datasets; (8) allowing citizens to post, rate, and work with datasets and web services; (9) a clear user interface; (10) standards for data, metadata, licenses, URIs, and exchange protocols; (11) integrate metadata schemas and federated controlled vocabularies; and (12) application programming interfaces for open data provision.

As (Maguire and Longley, 2005, p. 3) noted, "geoportals are World Wide Web gateways that organize content and services such as directories, search tools, community information, support resources, data and applications". The Geospatial One-Stop emerged as an easier, faster, and less expensive gateway for searching relevant geographic information sponsored by the US Federal Government (Yang et al., 2007). As (Goodchild et al., 2007, p. 250) pointed out, "humans have always exchanged geographic information, but the practice has grown exponentially in recent years with the popularization of the Internet and the Web, and with the growth of geographic information technologies. The arguments for sharing include scale economies in production and the desire to avoid duplication. The history of sharing can be viewed in a three-phase conceptual framework, from an early disorganized phase, through one centered on national governments as the primary suppliers of geographic information, to the contemporary somewhat chaotic network of producers and consumers". Many governments have been developing various programs to open data available via websites for public consumption such as the US, the UK, and Canada. These datasets typically contain records with spatial properties in democratizing public sector data and driving innovation (Arribas-Bel, 2014). Data.gov as a website launched in 2009 aims to enhance accessing the repository for federal government information and to ensure better accountability and transparency over 50 US government agencies with over 194,708 datasets (Hendler et al., 2012; Data.Gov, 2017).

In line with the spirit of crowdsourcing and citizen science, open data movement is that data should be legally and technically open to the scientific community, industry, and the public to use and republish. In other words, data should be provided in open machine-readable formats and readily located, along with the relevant metadata evaluating the reliability and quality of the data to promote increased data use and facilitate credibility determination. To advocate both transparency and innovation, open government data initiatives have been implemented in many countries from local to global level regarding accessibility, persistent identification, and long-term availability. The open data initiatives encourage peer production, interactivity, and user-generated innovation, which have stimulated the sharing and distribution of information across communities and disciplines. The underlying philosophy of open government or the theory of open source governance is that the interested citizens can access the documents of the government to facilitate effective public oversight and enable the direct involvement in the legislative process to promote openness, participation, and efficiency in government (Janssen et al., 2012). Civic hacking is utilizing government data to make governments more accountable through solving civic problems by those who care about their communities. For instance, Code for America is a nonprofit organization founded in 2009 to deal with the growing cross-sector gap in their effective use of technology and design for good (Wadhwa, 2011). Some civic hackers are employed by Code for America. National Day of Civic Hacking is a nationwide day of action where various developers come together to coordinate civic tech events dedicated to civic hacking (Johnson and Robinson, 2014). Transparency and participation through data integration and dissemination across domains and boundaries will facilitate collaboration among researchers, private sectors, and civilians leveraging their skills to help the society. (Gurstein, 2011) also examined the impact of such open data initiatives on the poor and marginalized and call for ensuring a wide opportunity for effective data use in the context of digital divide.

Papers with publicly available datasets usually have a higher citation rate and visibility than similar studies without available data, either by a direct data link or indirectly from cross-promotion (Piwowar and Vision, 2013). However, open data movement often faces the economic, legal, organizational, political, social, and technical challenges at either individual or institutional levels. Many researchers are still reluctant to share the original data used in their research to support the open access initiative, fearing the loss of credits, future publications, and competitive advantage, as well as the time to document and deal with questions for users (Fernandez, 2010; Sui, 2014). Their motivation for publication of datasets and intentions in doing so remain uncertain. Arguably, different person-based policies among stakeholders with various backgrounds and interests need to be developed to encourage sharing behavior in collaboration. Organizational support has been playing a substantial role in promoting researchers' intentions of sharing datasets through dealing with the heterogeneity of collaborators and the complexity of the data sharing process.

1.05.3 Open Source GIS

The Free and Open Source Software for Geospatial Conference has been playing a pivotal role in promoting the open science in software development. Open source GIS is gaining growing market shares in academia, business, and public administration. This recognition came at a time when many open source programming and scripting languages such as Python and R are starting to make major inroads in geospatial data production, analysis, and mapping (Ye and Rey, 2013). As a consequence, open source software development has been a crucial element in the GIS community's engagement with open GIS and the most well-developed aspect of open GIS (Rey, 2009). The availability and widespread use of codes and tools to support more robust data analysis will play a critical role in the adoption of new perspectives and ideas across the spatial sciences. The openness to scrutiny and challenge underlies the open source GIS movement through the release of the source code, which has subsequently influenced software functionality and support (Neteler and Mitasova, 2008). Users have the freedom to access, modify, and distribute the source code based on licensing agreements such as MPL, MIT, Apache, GPL, and BSD. Making source code both legally and technically open is the very first step of being promoted as a public good (Rey and Ye, 2010). In particular, scientists could benefit from the open source code, which would reduce code duplication and free up additional developer time to enhance the respective applications (Rey, 2009). Bonaccorsi and Rossi (2003) argued that "when programmers are allowed to work freely on the source code of a program, this will inevitably be improved because collaboration helps to correct errors and enables adaptation to different needs and hardware platforms". The credibility of research findings tends to be higher for papers with the available code. Third-party researchers might be more likely to adopt such papers as the foundation of additional research. In addition, coding repository platforms such as GitHub and BitBucket are making this open source tide stronger. Links from code to a paper might enhance the search frequencies of the paper because of accelerated awareness of the methods and findings.

The dramatic improvement in computer technology and the availability of large-volume geographically referenced data have enabled the spatial analytical tools to move from the fringes to central positions of methodological domains. By and large, however, many existing advanced spatial analysis methods are not in the open source context. The open source and free approach offer unprecedented opportunities and the most effective solution for developing software packages through attracting both users and developers. Instead of reinventing the wheel, we can study how the program works, to adapt it, and to redistribute copies including modifications from a number of popular alternatives. Anselin (2010) emphasized the role of the open source software movement in stimulating new development, transcending disciplinary boundaries, and broadening the community of developers and adopters. With accelerated development cycle, open source tools can give GIS users more flexibility to meet the user community needs that are only bound by our imaginations, which are aligned with more efficient and effective scientific progress. New theories and novel practices can thus be developed beyond narrowly defined disciplinary boundaries (Sui, 2014). Regarding open source efforts on spatial analysis, Arribas-Bel (2014) argued, "the traditional creativity that applied researchers (geographers, economists, etc.) have developed to measure and quantify urban phenomena in contexts where data were scarce is being given a whole new field of action". Sui (2014) also noted that a hybrid model integrating both open/free paradigm and proprietary practices (copyright and patent, IP stuff) would be the most realistic option and promising route to move GIS forward. Open source GIS can facilitate the interdisciplinary research due to "the collaborative norms involving positive spillover effects in building a community of scholars" (Rey, 2009; Ye et al., 2014).

During the past several decades, burgeoning efforts have been witnessed on the development and implementation of spatial statistical analysis packages, which continue to be an active area of research (Rey and Anselin, 2006; Anselin, 2010). The history of open source movement is much younger, but its impact on GIS world is impressive (Rey, 2009). As Rey (2009) commented, "a tenet of the free software (open source) movement is that because source code is fundamental to the development of the field of computer science, having freely available source code is a necessity for the innovation and progress of the field". The development of open source packages has been boosted. However, many duplicates and gaps in the methodological development have also been witnessed. The open source toolkit development is community-based with developers as well as casual and expert users located everywhere. Through the use of an online source code repository and mailing lists, users and developers can virtually communicate to review the existing code and develop new methods. However, Tsou and Smith (2011, p. 2) argued that "open source software is not well adopted in GIS education due to the lack of user-friendly guidance and the full integration of GIS learning resources". Some representative open source desktop GIS software packages include KOSMO, gvSIG, uDig, Quantum GIS (QGIS), Geographic Resource Analysis Support System (GRASS), and so on. KOSMO was implemented using the Java programming language based on the OpenJUMP platform and free code libraries. Developed by the European GIS community offering multiple language user interfaces, gvSIG is known for having a user-friendly interface, being able to access a wide range of vector and raster formats. Built upon IBM's Eclipse platform, uDig (user-friendly desktop Internet GIS) is an open source (EPL and BSD) desktop application framework. QGIS integrates with other open source GIS packages such as PostGIS, GRASS GIS, and MapServer, along with plugins being written in Python or C++. As a founding member of the Open Source Geospatial Foundation (OSGeo), GRASS offers comprehensive GIS functions for data management, image processing, cartography, spatial modeling, and visualization.

1.05.4 Practicing Open Source GIS

The study of economic inequality and convergence continues to attract enormous attention thus becoming a dynamic academic landscape where the interdisciplinary literature has evolved (Ye and Rey, 2013). This interest has been reflected in the analysis of spatial patterns of economic convergence and the temporal dynamics of geographical inequality. However, the literature studies of process analysis and form analysis are mainly separated because most methods are standalone without the sharing of the code. At the same time, the increasing availability of open space–time data has outpaced the development of space–time analytical techniques across social sciences. The methodological integration of space and time call for open data and open source methods, which will help narrow the gap between growth theories and their empirical testing. While the substantive focus of this case study is on open source computing of regional income dynamics, the issues examined are relevant to the development of a wide class of methods based on open science. This section suggests some novel exploratory approaches to compare spatial pattern and temporal trend of regional development. The cross-fertilization between domain science and open source computing is identified and illustrated.

There is a growing list of papers using local indicator of spatial autocorrelation (LISA) to measure the spatial structure of socioeconomic patters due to the following two reasons: the availability of open socioeconomic data across administrative units and the implementation of LISA indicators in software packages (Anselin, 1995). However, the release of spatial panel data calls for the extension of static spatial methods into the dynamic context. A LISA time path can be used to measure how economies covary over space and time in a regional context, giving insights to the debate about cooperative versus competitive regional growth. The challenging issue is that most domain users cannot handle the LISA time path due to the lack of programming skills.

Once the new space–time indicators are developed, an extensive set of inferential approaches is needed to evaluate their sampling distributions for comparative analysis between two regional systems. A tortuous LISA time path indicates that the focal economy and its average neighbor have instable convergence/divergence rates, while a frequent crossing suggests mixed convergence and divergence trends over time. With the growing awareness of the potential importance of the spatial dimension of economic structure, these space–time constructs can be implemented into empirical specifications to test the existence of poverty traps, convergence clubs, and spatial regimes. Two interesting questions arise from this analysis (the value hereafter refers to tortuous indicator or crossing ratio):

1. Is a LISA time path statistically more or less tortuous than what is expected if the path is randomly organized?
2. Is a LISA time path's crossing ratio statistically larger/smaller than what is expected if the path is randomly organized?

Both LISA time path and various simulation procedures can be modified based on codes from open source packages STARTS and PySal (Rey, 2009, 2014). It is expected that LISA coordinates are independent from each other; however, an individual region's economic growth at one time point relates to its history and its neighbors' temporal economic dynamics. An alternative to the aforementioned methods is to employ a Monte Carlo simulation approach and thus circumvent the assumption of independence that causes inferential problems. The presence of space–time effects needs to be considered when examining the distributional properties of a LISA time path indicator. Three sets of permutation approaches are suggested to test the independence of space, time, and trend through python code implementation.

The spatial independence test answers the following questions: Can the observed value (or the differences between two observed values) be used to reject the null hypothesis of spatial randomness? The temporal independence test answers the following question: Can the observed value (or the differences between two observed values) be used to reject the null hypothesis of temporal randomness? The trend independence test answers the following question: Can the observed value (or the difference between two observed values) be used to reject the null hypothesis of trend randomness?

Figs. 1–3 show the effects of these three independence tests, the top left view is a LISA time path of region i from Time 1 to Time 10. In Fig. 1, through random permutations on the spatial coordinates of all regions, the other three graphs show three different LISA time paths with different groups of points (LISA coordinates). In Fig. 2, through randomly relabeling time stamps, the other three graphs show different LISA time paths based on the same group of points. In Fig. 3, through randomly normalization the path segments to follow a normal distribution, the other three graphs show different LISA time paths that retain the trends of the original path.

We test LISA path significance using tortuosity and crossing ratio using provincial GDP data in mainland China from 1998 to 2008, consisting of 31 provinces in total. This work uses the k-nearest neighbors method to construct the spatial matrix, and default k is 4. The related code is adopted from PySal (http://pysal.readthedocs.io/en/latest/users/tutorials/weights.html). We compute the observations $\boldsymbol{o}_t$ and their lag values $\boldsymbol{l}_t$ for each time points. Both values are standardized as z-scores. Construct a dictionary $\boldsymbol{D}$ that uses the time point t as key, and a matrix comprised two rows transposed from $\boldsymbol{o}_t$ *and* $\boldsymbol{l}_t$. We then construct the observed LISA path with the values extracted from the dictionary $\boldsymbol{D}$. A LISA time path P_i for a given province i consists of several spatial coordinates with their temporal labeling, represented as:

$$P_i = \{(x_k,\ y_k,\ t_k)\},\ 0 < k < n + 1$$

The first two elements x_k and y_k in the tuple are the attribute value and its spatial lag value, respectively, of the province i, while t_k is the k_{th} year when the value is measured. Finally, we obtain a set of paths Sp for all provinces. For each path, Pi in the path set Sp, two indexes are integrated in the following simulation procedures: tortuosity and crossing ratio. Tortuosity is represented as:

$$T_p = distance((x_1, y_1), (x_n, y_n))/len(P)$$

where (x_1,y_1) and (x_n,y_n) are the head and tail spatial coordinates of the LISA path P, and $len(P)$ measures the total length of all the segments of the path. The value of tortuosity is in the range [0, 1], with 0 representing higher degree of path tortuosity, and 1 for the path to be completely straight spatially, stretching in a stable direction.

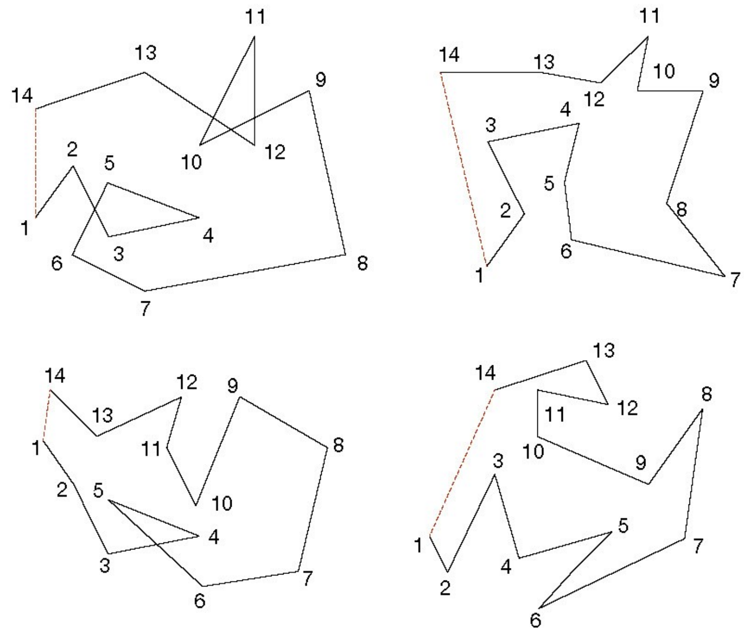

Fig. 1 Spatial independence test.

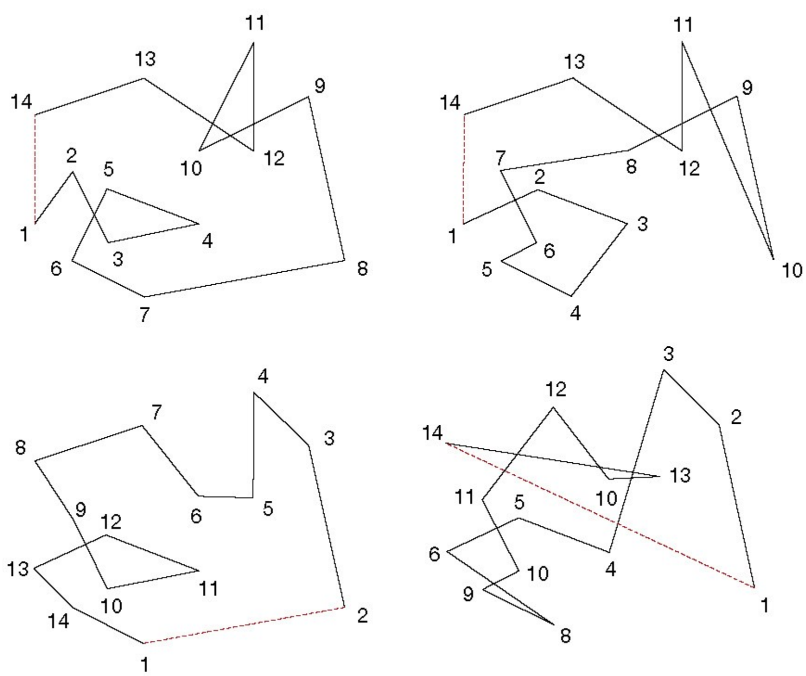

Fig. 2 Temporal independence test.

The crossing ratio is represented as:

$$c_p = 2 * ic_p \Big/ \left(n_p{}^2 - n_p\right)$$

where ic_p represents the self-intersection count of the LISA path p, and n_p represents the number of points in p. The self-intersection count is calculated by checking whether segments of this path P intersect with each other. The value of crossing ratio also lies in the range [0, 1], with 0 representing no crossing, which indicates that the path is highly stable, and 1 for another extreme case where all the segments in the path happen to intersect with each other, indicating highly unstable evolvements of the path over time.

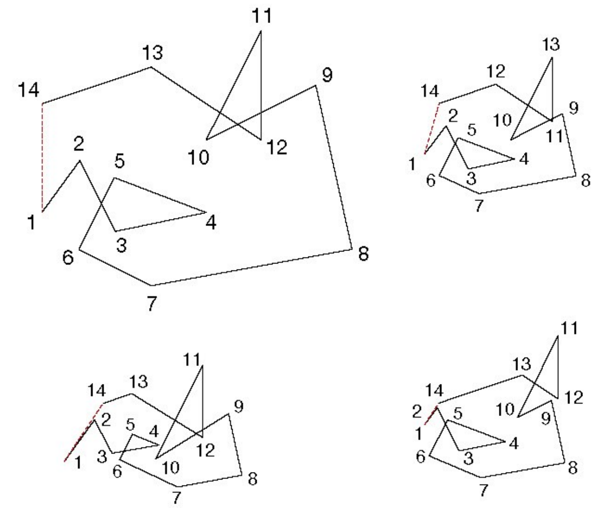

Fig. 3 Trend independence test.

We permute the original data 999 times using one of the following three modes: spatial independence, temporal independence, and trend independence.

(i) Spatial independence test

The spatial independence acts on all the provinces. For each time point t, randomly permutate the spatial coordinates for all the regions. This process rearranges the LISA and the lag values among the provinces, and a new LISA time path is constructed for the given region.

(ii) Temporal independence test

The temporal independence test uses the data solely from the observed LISA path. The test randomly permutates the temporal order of the spatial coordinates of the given region and form a new LISA time path.

(iii) Trend independence test

Similar to the temporal independence test, the trend independence test takes only the data from the observed LISA path. The test starts with breaking the observed LISA path into a set of vectors. These vectors are then normalized to follow a normal distribution centered at zero and with unit length as standard deviation while preserving their directions. A trend list is thus formed. The new LISA time path is generated by first randomly picking a starting coordinate and then uses the trend list to construct the whole path.

The values calculated in the simulation process are then ordered, and the pseudo significance level is then computed. This simply sorts the empirically generated 999 values and then develops a pseudo significance level by calculating the share of the empirical values that are higher than the actual value.

The results show that spatial independence and temporal independence tests show stronger significance level than trend independence test. The ranks of the tortuosity and crossing ratio in each test are not much correlated but are relatively stable respectively across three tests. Under the hypothesis of spatial or temporal independence, all provinces are significantly tortuous than expected, except three provinces (Jiangxi, Jilin, and Shanghai). However, none of the provinces shows significant crossing frequencies. In other words, there are no obvious mixed convergence and divergence trends over time.

1.05.5 Summary

Spatial turn in many socioeconomic theories has been noted in a vast field, encompassing both social and physical phenomena (Krugman, 1999; Goodchild et al., 2000; Batty, 2012). The fast growth in socioeconomic dynamics analysis is increasingly seen as attributable to the availability of space–time datasets (Rey and Ye, 2010; Ye et al., 2016). Rigorous space–time analysis and modeling open up a rich empirical context for scientific research and policy interventions. To help scholars and stakeholders deal with the challenges and issues, methodologies and best practice guidelines are needed at both international and national or local level. Making data and source code available for both replication and continuing research will have far-reaching and broader impacts in both GIS and domain communities, highlighting the growing recognition of the role of geography in interdisciplinary research (Karnatak et al., 2012). Such improved discoverability is beneficial for both investigators and the science community as a whole. Open source GIS research has received increased attention and will lead to the revolutionary advances in individual and collective decision-making processes (Sui et al., 2012). The goal of this article is to make a modest effort to synthesize an agenda surrounding and hopefully to stimulate further discussions that promote open GIS as the driving force to guide the development of

GIS at a finer scale (Sui, 2014). To gain momentum under the general umbrella of big data and new data, GIS should fully embrace the vision of an open data and open source GIS paradigm to enhance government efficiency and to improve citizen services (Wright, 2012). Although there are academic, legal, social/political, and environmental impediments for the practice, open GIS will provide numerous technology-driven, application-led, science-inspired, and education-focused opportunities (Sui, 2014).

Methods developed in the mainstream spatial science disciplines need to be progressed with more attention paid to the potential reuse of data and codes. Though a growing list of research papers have highlighted the increasing awareness of spatio-temporal thinking and action, the gap has been widening between a small group of method developers and users. Hence, a crucial step is to develop the dialog between computational scientists and domain users, seeking the cross-fertilization between the two fast-growing communities. As Rey (2009) suggested, "increased adoption of open source practices in spatial analysis can enhance the development of the next generation of tools and the wider practice of scientific research and education". The methods are built in open source environments and thus easily extensible and customizable. Hence, open source project can promote collaboration among researchers who want to improve current functions or add extensions to address specific research questions.

Acknowledgments

This research was funded by the National Science Foundation (1416509, 1535031, 1637242).

References

Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89 (1), 3–25.

Anselin, L., 1995. Local indicators of spatial association—LISA. Geographical Analysis 27 (2), 93–115.

Arribas-Bel, D., 2014. Accidental, open and everywhere: Emerging data sources for the understanding of cities. Applied Geography 49, 45–53.

Auer, S. R., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z. (2007). DBpedia: A nucleus for a web of open data. In: *The semantic web* (Lecture notes in computer science, vol. 4825), p. 722. 10.1007/978-3-540-76298-0_52.

Batty, M., 2012. Smart cities, big data. Environment and Planning-Part B 39 (2), 191.

Bonaccorsi, A., Rossi, C., 2003. Why open source software can succeed. Research Policy 32 (7), 1243–1258.

Chignard, S. (2013). A brief history of open data. *ParisTech Review*, 29 March.

Data.Gov (2017) http://www.data.gov/ (accessed 20 February 2017).

Fernandez, R. (2010). Barriers to open science: From big business to Watson and Crick. http://opensource.com/business/10/8/barriers-open-science-big-business-watson-and-crick (accessed 16 April 2017).

FGDC.Gov (2017). https://www.fgdc.gov/metadata (accessed 20 February 2017).

Goodchild, M.F., 2013. The quality of big (geo) data. Dialogues in Human Geography 3 (3), 280–284.

Goodchild, M.F., Anselin, L., Appelbaum, R.P., Harthorn, B.H., 2000. Toward spatially integrated social science. International Regional Science Review 23 (2), 139–159.

Goodchild, M.F., Fu, P., Rich, P., 2007. Sharing geographic information: An assessment of the Geospatial One-Stop. Annals of the Association of American Geographers 97 (2), 250–266.

Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday* **16**(2). (accessed 16 April 2017).

Hendler, J., Holm, J., Musialek, C., Thomas, G., 2012. US government linked open data: Semantic. Data.Gov. IEEE Intelligent Systems 27 (3), 25–31.

Janssen, M., Charalabidis, Y., Zuiderwijk, A., 2012. Benefits, adoption barriers and myths of open data and open government. Information Systems Management 29 (4), 258–268.

Jiang, B., 2011. Making GIScience research more open access. International Journal of Geographical Information Science 25 (8), 1217–1220.

Johnson, P., Robinson, P., 2014. Civic hackathons: Innovation, procurement, or civic engagement? Review of Policy Research 31 (4), 349–357.

Kamatak, H., Shukla, R., Sharma, V., Murthy, Y., Bhanumurthy, V., 2012. Spatial mashup technology and real time data integration in geo-web application using open source GIS – A case study for disaster management. Geocarto International 27 (6), 499–514.

Krugman, P., 1999. The role of geography in development. International Regional Science Review 22 (2), 142–161.

Maguire, D.J., Longley, P.A., 2005. The emergence of geoportals and their role in spatial data infrastructures. Computers, Environment and Urban Systems 29 (1), 3–14.

Neteler, M., Mitasova, H., 2008. Open source GIS: A GRASS GIS approach, 3rd ed. Springer, Berlin.

Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R., Muro-Medrano, P.R., 2004. Metadata standard interoperability: Application in the geographic information domain. Computers, Environment and Urban Systems 28 (6), 611–634.

Piwowar, H.A., Vision, T.J., 2013. Data reuse and the open data citation advantage. Peerj 1 (175), 1–25. http://dx.doi.org/10.7717/peerj.175.

Rey, S., 2009. Show me the code: Spatial analysis and open source. Journal of Geographical Systems 11 (2), 191–207.

Rey, S.J., 2014. Open regional science. The Annals of Regional Science 52 (3), 825–837.

Rey, S.J., Anselin, L., 2006. Recent advances in software for spatial analysis in the social sciences. Geographical Analysis 38 (1), 1–4.

Rey, S., and Ye, X. (2010). Comparative spatial dynamics of regional systems. In: Páez, A., Le Gallo, J., Buliung, R. & Dall'Erba, S. (eds.) *Progress in spatial analysis: Theory, computation, and thematic applications*, pp. 441–464. Springer: New York City.

Stodden, V., 2009. The legal framework for reproducible research in the sciences: Licensing and copyright. IEEE Computing in Science and Engineering 11 (1), 35–40.

Sui, D., 2014. Opportunities and impediments for open GIS. Transactions in GIS 18 (1), 1–24.

Sui, D., Elwood, S., Goodchild, M., 2012. Crowdsourcing geographic knowledge: Volunteered geographic information in theory and practice. Springer, Berlin.

Susha, I., Zuiderwijk, A., Charalabidis, Y., Parycek, P., Janssen, M., 2015. Critical factors for open data publication and use: A comparison of city-level, regional, and transnational cases. JeDEM-eJournal of eDemocracy and Open Government 7 (2), 94–115.

Tsou, M. H., & Smith, J. (2011). *Free and open source software for GIS education*. White paper written with the support from the National Geospatial Technology Center of Excellence (GeoTech Center, http://www.geotechcenter.org/).

Wadhwa, V. (2011). Code for America: An elegant solution for government IT problems. *The Washington Post.*

Warf, B., Arias, S. (Eds.), 2008. The spatial turn: Interdisciplinary perspectives. Routledge, London, UK.

Wright, D. (2012). Big data, GIS, and the academic community. http://blogs.esri.com/esri/esri-insider/2012/10/03/big-data-gis-and-the-academic-community/#more-1311.(accessed 15 April 2017).

Yang, P., Evans, J., Cole, M., Marley, S., Alameh, N., Bambacus, M., 2007. The emerging concepts and applications of the spatial web portal. Photogrammetric Engineering & Remote Sensing 73 (6), 691–698.
Ye, X., Rey, S.J., 2013. A framework for exploratory space-time analysis of economic data. Annals of Regional Science 50 (1), 315–339.
Ye, X., She, B., Wu, L., Zhu, X., Cheng, Y., 2014. An open source toolkit for identifying comparative space-time research questions. Chinese Geographical Science 24 (3), 348–361.
Ye, X., She, B., Zhao, H., Zhou, X., 2016. A Taxonomic Analysis of Perspectives in Generating Space-Time Research Questions in Environmental Sciences. International Journal of Applied Geospatial Research doi. http://dx.doi.org/10.4018/IJAGR.2016040104.
Ye, X., Yu, J., Wu, L., Li, S., Li, J., 2015. Open source point process modeling of earthquake. In: Geo-informatics in resource management and sustainable ecosystem. Springer, Berlin and Heidelberg, pp. 548–557.