

Geographical Analysis: Reflections of a Recovering Editor

Sergio Rey

December 18, 2018

1 Introduction

In this paper I take a retrospective and prospective look at Geographical Analysis (GA) the journal, as well as the broader academic field. I first set the context for my comments by giving an overview of my history from a personal perspective and how my career path has intersected with both the journal and the field of spatial analysis. This is followed by a section that covers three external trends that I think will form the backdrop for the future evolution of spatial analysis and Geographical Analysis the Journal. These include the rise of artificial intelligence, machine learning, and data science as disciplines as well as some structural changes in the organization of science and appearance of competing models for that organization. The third large-scale trend that we will intersect with has to do with the democratization of spatial analysis and of geospatial data. After covering these trends, I then move onto areas that I think have been under-studied but are likely to be major areas for future research in spatial analysis. These concern multi-scalar integration and the problem of endogenous spatial units. In the final section of the paper I identify future opportunities for the journal *Geographical Analysis*.

2 History

To provide some context, I first discuss my history and association with the journal and spatial analysis in general. I was a graduate student at UCSB in geography in the early 1990s, where I obtained both my masters and doctoral degrees. Looking back now, I realize just how fortunate I was to do graduate studies there, as it was a period of substantial growth for the department, and one when GIScience was coming to the fore as a field.

The journal and the department seemed to be joined at the hip back then. GA had been positioned as the premier outlet for advanced theoretical research in quantitative geography. The department, had by that time, established itself as one of the world's leading spatial analysis centers. Two of the former editors of the journal, Reg Golledge and Mike Goodchild, were faculty at UCSB during this time. It is also with some personal pride to note that the department would also produce another future editor of GA: Alan Murray.

One of my first papers appeared in GA, and I still remember the feeling of excitement in seeing my name in print in the table of contents. I don't think I am alone in my high regard for the journal as I've heard from many colleagues that they view GA as the flagship journal for the best in quantitative spatial analysis. The high regard can sometimes reflect strongly held personal interactions with the journal, and it is here where I want to reflect on a painful but important piece of history.

I have always felt that the health of a journal can be measured in the strength of the community that is built around it. Scholars associate themselves with a journal through publishing their best work in the journal, contributing as editorial board members and referees, as well as serving on the executive staff or as editor. I was honored to join the editorial board in 1994, and very much enjoyed the opportunity to read some of the latest in spatial analysis and provide suggestions and referee reports for the editor.

Because I highly value community, I resigned from the editorial board in 1998 over what I felt was a broken governance model that was treating members of the community poorly. The reaction was swift. I received numerous admonishments from senior colleagues for the damage they feared such a move would cause to the journal. Yet, other senior colleagues quickly followed my action and also resigned. This turmoil was unfortunate, however, I felt I did the right thing by resigning.

With hindsight, I have learned two lessons from this episode. First, time can indeed heal wounds. Second, the high regard members of our community felt for the journal, and honest disagreements about how the journal should function are actually healthy signals. As evidence of the healing power of time, in 2013, I was approached by the executive board to consider becoming the editor. I report this episode because I think it reflects the maturity of GA community in that disagreements can arise when people feel strongly about an issue. How they resolve those disagreements says a lot about the health of the community.

I served a three year term at GA (2014-2017), and had I not preceeded this with a 16 year term at the International Regional Science Review (IRSR),

I may have sought a second term. I mention this here not to self-promote but, rather, to say that I was struck by some major differences between my experience as editor on these two journals. At GA, I experienced a much more engaged editorial board. Numerous colleagues offered constructive criticisms and support, beyond referee reports, to me during my term as editor. These reflected a genuine concern for GA and a shared sense of responsibility. This is not a criticism of my colleagues who served on the IRSR board over my tenure - it is more a recognition of one of the distinguishing characteristics of GA.

3 Emerging Trends

Although prognostications about the future are always dangerous, I see three prominent emerging trends that, I think, will shape the future of Geographical Analysis the journal and the field: The rise of machine learning, artificial intelligence and data science; the emergence of the competing models of geospatial analysis; the democratization of spatial data and analysis.

3.1 The Rise of Machine Learning, Artificial Intelligence, and Data Science

The turn of the century has witnessed an explosion of interest in the fields of machine learning, artificial intelligence, and data science. This has generated an abundance of applications of these technologies and methods across a wide array of substantive domains. More importantly, it has raised fundamental epistemological issues for traditional disciplines - geography and spatial analysis have been somewhat isolated from these events, although I expect this to intercede more fully in the coming years.

One main issue that sits before geographical analysis is in the realm of the Big Data era. Here the often quoted phrase is that data is the new oil. Companies like Google and Facebook and Twitter have mined massive amounts of data sources, and so-called data lakes, to develop services such as search and social graphs that have transformed our world. It is clearly true that these new forms of data have been critical to the success of these companies. But I would argue that if data are the new oil, and I believe this is true, that analytics are the new refineries. And it is here where I see fantastic opportunities for spatial analysis.

The widespread availability of geospatial web-based technologies has lowered the cost of entry to anyone wanting to develop mashups surrounding mapping and other types of visualizations. As a result, there has been an

explosion of such applications. At the same time, however, one gets the feeling that we may be hitting diminishing returns to these efforts. Visualization leads to natural questions about what processes are responsible for the patterns that these innovative visualizations are uncovering. Answering those questions is the bread-and-butter of quantitative geographic analysis. However, engaging with the world of web-mashups for geospatial data is something that we are not capturing or realizing at this moment as a community. Analytics are the mechanism to turn visualizations into information and knowledge, and spatial analytics in particular are clearly the next generation of refineries in the big data realm.

Computational thinking [1] is a second area that the new data science challenge offers to move spatial analysis forward. Computational thinking gives us new ways to engage students with advanced concepts. For example, by beginning with the final solution to a problem, say a location allocation problem or spatial search problem, and only then unveiling the solution through the details of the algorithm, students feel a sense of excitement about why the underlying fundamentals are important, and because of that recognition, engage with those fundamentals in a deeper sense, thus lowering the bar to advancing their skill sets.

Adopting computational thinking as a pedagogical strategy helps students acquire the mindset that when facing a challenging problem that appears daunting, they can re-frame that to see the learning process in this particular case has a slow feedback loop. That is, they recognize they will eventually learn these concepts, it just takes longer when dealing with advanced concepts relative to easier concepts. That recognition lowers the frustration level as the expectations are aligned with reality. This is a subtle, but important, shift in the way we present material and encourage students to become lifelong learners.

To facilitate computational thinking there have been a number of fantastic innovations in scientific computing that we can lean on in reshaping the teaching of spatial analysis. Groups such as Software Carpentry and Data Carpentry ¹ have adopted the bootcamp model to give faculty, students, and staff of educational institutions primers on scientific scripting and, in the case of the data carpentries, these are tailored to particular domains such as bioinformatics, atmospheric, and ocean sciences. These are often very successful in giving a shot in the arm to the participant in getting them so called over the computational hump.

As a community, academic spatial analysis has been slow to engage with

¹<https://carpentries.org>

these pedagogical transformations. A cursory examination of GA contents shows scant coverage of the teaching of quantitative spatial analysis. I find this troubling for two reasons. First, the lack of attention to spatial analysis pedagogy in GA reflects the historical realities that have prioritized research over teaching in the field of spatial analysis. While understandable, this is really a failure on the journal's part. By neglecting pedagogy, GA removed itself from the training of the future generations of theoretical quantitative geographers.

The second reason I find this neglect troubling is that academic spatial analysis is missing an opportunity to contribute to, and shape, the widespread adoption of geospatial methods out in the wild. Increasingly at computationally oriented conferences I attend, and from reading blogs high profile data scientists, I see the rise of what I will term as "Code Hubris". By this, I mean the attitude that by learning basic principles of software development and generic data science, one becomes an authority on the applications of these methods to domains outside one's training. The danger here is that the hard learned lessons about the special nature of geospatial data that academic quantitative geographers have to offer are being ignored by the newly minted data scientists as they seek to conquer the world. As a result, their data science is repeating mistakes we made decades ago [2].

A major question that the journal will have to address is how to position itself in the new big Data and machine learning error. Its historical roots in theory [3] could be seen as either a burden or a strength in this regard. The first perspective reflects the dominant thinking in the machine learning community in that the goal is prediction and not understanding the causal processes underlying the patterns that have been recognized by the latest machine learning algorithm. Any machine learning researcher worth their salt is open about this and does not over-extend by claiming that these approaches yield causal insights. Some, however, have gone beyond this measured perspective to claim that theory is no longer needed as the growing power of machine learning prediction will be sufficient[4].

I think this is overblown and misplaced for two reasons. First, the tension between pattern and process is an artificial one as any geographer knows. Indeed, one cannot separate consideration of pattern from questions about processes that may be at work under the hood. Second, seeing predictive methods as orthogonal to causal inquiry shows a lack of imagination. I see the two sets of methods as complements, and by tapping into that relationship we can be on a path to better science. And, by better science here I mean an accelerated innovation cycle where we go from pattern discovery to posit causal mechanisms in a much quicker cadence than we would if we

have separated the researchers working on exploratory methods from those working on process. The new spirit of computational thinking is blurring the distinction between the traditional spheres of investigation.

3.2 Cathedrals or Bazaars

A related trend I can foresee is a competition between different models for organizing the field of spatial analysis. Here I think Eric Raymond's Cathedral and the Bazaar [5] offers the appropriate metaphors for contrasting what I think are competing alternatives for the future spatial analysis. The Cathedral model is reflected in large companies controlling the distribution of spatial analysis software. In the sense that many researchers turn to esri software and think that it is the final arbiter of what to use when carrying out spatial analysis. This has advantages in terms of having a well-recognized standard representation of spatial analysis software. But, it has costs as well. The primary costs are in the centralization of the development model and vendor lock-in.

The alternative is the bazaar model in which research is decentralized and different teams collaborate and/or compete to develop the next generation of spatial analytical tools and software. In the bazaar model, no central authority exists. Instead, structure emerges organically through institutions and organizations working collaboratively on the code base. GitHub² is particularly interesting as it has become the place where developers post open source code to tap into the distributive development model that git affords. This open source model has advantages in terms of tapping into network effects and having the coordination implicitly built into the system.

Moving forward, it is not clear which of these two models will come to dominate spatial analysis in the future. There are some arguments that open source has actually lost the battle already [6] in that large companies such as Google have co-opted the open-source model, and in clever ways centralized the rewards of that model to produce profitable products. For example, their web search uses Python and related packages under the hood and are produced as web services. Google has been criticized for being a consumer of open source tools and reaping vastly more benefits than it has returned to the community. Whether these are fair criticisms is subject to debate. On the benefits side I would be remiss, however, not to mention the contributions Google has made to spatial analysis. Through their summer of code thousands of student developers have been supported on open source

²<https://github.com>

projects, many of them in the spatial analysis domain (e.g., PySAL, osgeo).

Google is not the only company using open source to support its business model. Agile companies such as Mapbox and Carto have arisen which have taken a more collaborative stance with open source projects to build new products around open source technologies. The question is whether spatial analysis software becomes controlled by a geospatial analytics industrial complex (i.e., the Cathedral model) or will take the form of the more distributed organization (i.e., the bazaar model).

3.3 Democratization of Geospatial Data and Analysis

A third development in future of geographical analysis will be in the area of the democratization of analysis and data. This is closely related to the competition between the different organizational models touched on previously. In large part, these developments are driven by the open source movement where the source code underlying a spatial analytical method is made publicly available for the reading. The other aspect of “open” is that the source code is available for modification and extension. Both of these freedoms are critical to fueling growth in the number of people who can engage with spatial analysis methods and, ultimately, the increase in the advances in the form of new methods.

Facilitating that openness this will be increasingly important to maintain and enhance these channels of innovation [7]. As I touched on previously, it is not at all clear if the open source model as the dominant paradigm will be a foregone conclusion as there are challenges arising on the horizon that need to be grappled with.

The democratization of open data is an equally important development to the democratization of analysis. Currently there is a developing *geospatial data provision industrial complex* and that is doing quite well. Companies [8] as well as research groups [9, 10] have taken publicly-available data in the form of census and American Community Survey data, in the US case, and added value through easy to deliver front-ends that provide one-stop-shopping to researchers looking for accessible data. These companies and groups provide valuable services to the research community and have supported thousands of empirical studies.

While impressive in their impact and scope of use, this model is not without problems. A number of prominent data providers charge subscription, or access, fees for their data. This seems logical given the investment in time and effort put into building and maintaining the services. However, it does present a constraint for resource strapped individual scholars. The business

model that has arisen to relax this constraint is provide subscription fees to libraries which transfers the cost away from the individual user to the institution.

There is a second constraint in these use of these types of data services. Often the licensing agreement places restrictions on how the data, once accessed, can be used. Researchers may not redistribute the data without prior permission of these providers. This is a serious, but under-recognized, constraint scientific innovation. It prevents the further community enhancements to the databases.

I would argue that relaxing the restriction on downstream redistribution of the data would not necessarily prevent future resources from coming into the project [11]. An analogy here would be to look at the grandfather Linux distributions such as Debian³ which is released for free under open source licenses. Entire companies, such as Canonical⁴, have been built around packaging that software and providing services for its application. And this has not dampened the growth and expansion of the original Debian project. Mechanisms to support the project have included consulting fees by private company who want specialized applications, and support from governments who see the value in what the grandparent distribution is providing. Finally, contributions of members in the community who also value the importance and availability of the original distributions should not be ignored.

The type of model used by leading data providers in the geospatial domain can enhance the impact that their efforts are having on research productivity as well as maintain the financial viability of those projects. The trade-off between the benefits that the data providers offer to the community versus the constraints they place on downstream redistribution of the data represent a trade-off between data use and data innovation. A subtle, but important tweak of the system could increase the importance of innovation flowing from these projects by rethinking of the business models that underlie these data provision activities.

A final aspect to the democratization of both geospatial analysis methods and data relates to the motivation for engagement of individuals with these activities. Historically, it has been notoriously difficult to get funding from federal agencies for both open source tools development and their provision to the community. This is beginning to change as we have recently seen prominent examples of major investments by the US National Science Foundation in both of these areas. Here we have the rare instance

³<http://debian.org>

⁴<http://www.canonical.com>

of the government actually innovating and leading the charge and academia playing catch-up. By catch-up, I mean these types of contributions (data provision and scientific software development) have historically been undervalued in the formal evaluation of academic researchers [12], but are now given consideration during these evaluations.

4 Opportunities

Given these emerging trends, I can see opportunities for both the field of geographical analysis as well as for Geographical Analysis the journal.

4.1 Opportunities for Geographical Analysis: The Field

There are two areas of research that I feel are important future opportunities for spatial analysis to contribute significant advances. The first has to do with what I call multiscale integration, and the second concerns the problem of endogenous spatial units.

Multiscale Integration: Geographical analysis is fragmented methodologically in many ways but, in particular, from a scalar perspective. By this I mean we have excellent work being done on micro-scale models. These include agent-based models, cellular automata, and microsimulation models. Similarly, there are researchers who investigate more macro oriented models such as integrated multi-regional econometric and input-output models, and the wide scale use of spatial econometrics models which are typically focused on areal units as the organizational frame. We might also consider work at the meso-scale where the grouping observations from micro perspective to identify clusters as constellations of observations.

So at least we have individuals doing innovative work at these different spatial scales. However, if you attend our meetings, or read through our literature, there is precious little cross fertilization between these three groups of studies. This is unfortunate because I think it misses the opportunity to strengthen these areas of study.

There is a pressing need for integration across the scales: or what I refer to as multi-scale integration. Such an integration will lead to better spatial science in two regards. First, the comprehensive modeling frameworks that will arise from addressing integration will have much wider scope in the types of problems they can address relative to the piecemeal application of the individual-scale models. Second, even if one wants to remain in the domain of a particular scale, viewing the associated modeling frameworks for that scale from the perspective of a larger integrative context will enhance

the development of the individual components. By placing their work in a broader framework, researchers will be forced to consider more dimensions of their problems. Simply put, what was an exogenous variable when looked at from the isolation of a microscale perspective may become an endogenous variable from an integrated perspective. The same could be said of the meso and macro scales.

When one adopts an integrated multiscalar perspective, accuracy concepts also become expanded. Rather than focusing on the partitive accuracy of a particular scale-specific modeling framework, the concern shifts towards holistic accuracy: the accuracy of the integrated multi-scalar framework. Issues of error propagation and reverberation across components at different scales become central research topics.

Endogenous spatial units: The second promising area for advances in spatial analytical research has to do with the problem of endogenous boundaries. By and large, our current modeling frameworks in spatial analysis take space and its representation as fixed in a container sense, and we study how attribute distributions evolve over those fix units. This is clearly seen in the work on regional labor markets as well as housing markets where administrative boundaries are often used to record the data under study with the assumption that those administrative boundaries coincide with the boundaries of the actual market process. In practice, there is good reason to suspect these assumptions are restrictive. The notions of growth and change suggest that it is not only the magnitude of say the number of people in the labor market that changes over time but also the spatial extent of the market. Were this true, the assumption of fixed administrative boundaries as proxies for process boundaries is clearly violated.

Similarly, in the work on neighborhood effects and neighborhood dynamics, often census tracts and other administrative units are used to define neighborhoods, either explicitly as the neighborhood unit, or by aggregation to combine tracts that have similar socioeconomic composition and geographical connectivity. Once these neighborhood boundaries are defined, the focus shifts to the socioeconomic composition of the individuals and households who reside within the boundaries and how that composition may fluctuate over time. But again, the boundaries themselves are fixed and do not evolve in any true sense. What can evolve here, just to be explicit, is that the neighborhood labels applied to a given tract as the label for the neighborhood type it is assigned. The boundaries of the primitive units are not allowed to change, however.

Endogenous boundaries are thus an important area for future research. We have good reason to suspect that neighborhoods change not just in their

composition but also in spatial extent, and the same can be said of housing and labor markets. The challenge is how to incorporate boundary change into our analytical frameworks. At present we lack any good alternatives to turn to. But we might be able to draw inspiration from our colleagues in the physical sciences who model things such as storms and fires as systems that have attribute compositional dynamics but also spatial dynamics [13].

A little closer to home for quantitative economic geographers, we might think about the problem of say the diffusion of regional recessions as an analog to a storm. If we borrow modeling concepts from climatology and marry them with new types of data in the form of geo-social media, we may be able to allow for endogenous modeling of boundaries. Addressing the issue of endogenous boundaries can also intersect with the work on multi-scalar integration identified above. We can start to think about having micro level models based on geotagged social media data tweeting on the emergence of a recession and apply clustering algorithms to that real-time data to identify areas where similar sentiments about being in recession or not are expressed, and then use the resulting spatial footprint of the clusters to define macro regions. At a higher level, we could have a second-order model that handles the probability of being in or out of recession for the regions. Over time we iterate this model and include a feedback mechanism where the geosocial media tweets are embedded in a broader regional context from the previous period's solution. This can also be used to extend the clustering for the future periods and allow for explicit bottom-up and top-down linkages between the micro and macro scales.

4.2 Opportunities for Geographical Analysis: The Journal

Given the challenges facing the future of spatial analysis as well as the opportunities these challenges can represent it is interesting to think about what the GA as a journal can do to situate itself in the future. There are tensions between deciding whether to rebrand the journal and modernize in terms of the trends happening in the data science, machine learning, and artificial intelligence era and their articulation in spatial analysis. On the other hand, the journal has a well-established brand that has been hard-earned through the efforts of dedicated editors, and editorial board members over decades. That branding is important as reputations of journals are a key currency in the academic labor market. The rankings of journals matter for the evaluation of promotion cases and, as we've seen, that process gives scholars guidance and incentive to decide how to spend their research dollars on in terms of time and commitment.

Journals thus can have a key influence in shaping direction of a field. However that capability is a function of the currency of the journal. GA's brand is in its by-line "A theoretical journal in quantitative geographical analysis". Theory has been the distinguishing characteristic of the journal relative to the other major geographical outlets. The future rebranding of the journal will be tied to how theory is used in the new data science era. We have seen that theory offers some opportunities for engagement with machine learning and artificial intelligence, that I think to date have been ignored at best, or, at worst, dismissed as irrelevant. If GA were to serve as the forum for investigating the new role of theory and geographical analysis in the modern era I think it would do a major service for the community.

Branding, or rebranding, the journal will not be the only consideration GA faces in the future. Scientific publishing is undergoing fundamental changes in both the underlying technology used to produce scientific articles, as well as in the organization of the market and the engagement of scholars. Within spatial science we have been slow to tap into the opportunities that these technological and organizational shifts offer. The dead tree journal article is very likely to become a relic of the past as new forms of digital publication become viable. Innovations such as the Jupyter notebooks⁵ allow for powerful dissemination mechanisms where the narrative text of an article can be interwoven with the source code, visualisations, and other forms of media to present a much more comprehensive and reproducible depiction of the research effort.

Moreover, by tapping into the technology of software code repositories, the research artifact can become a living document in the sense that the source code is fully available to the researchers who want to build on the contribution as well as extended. There are amazing opportunities here to hard-wire attribution networks and build a genuine democratic approach towards science.

There have been some innovative journals in GIScience that have started to tap into new models of production. The Journal of Spatial Information Science is fully Open Access and makes its content available gratis to the world. Its status is maintained through a editorial board of highly-regarded GIScientists. Interestingly, many of the members of these boards also serve on the editorial boards of commercial journals, so clearly the two models can co-exist as far as the individuals are concerned. GA may want to consider these alternative models, in terms of the structure and content, and how the journal is delivered.

⁵<https://jupyter.org>

Having served for one term as editor of GA I, know how challenging it can be to think about, let alone try to implement, some of these structural changes on top of the normal duties facing an editor. Upon further reflection I don't think it is rational to put this all on the shoulders of the editor, but really should be a distributed effort that engages all contributors to the journal including authors, referees, and, hopefully, innovative scientific publishers. These are major challenges but I think if we as a community can rise to meet them, we will generate a better spatial science for it.

References

- [1] Lorena Barba. Computational thinking: I do not think it means what you think it means. <https://bids.berkeley.edu/news/computational-thinking-i-do-not-think-it-means-what-you-think-it-means>, March 2016.
- [2] Dani Arribas-Bel and Jon Reades. Geography and computers: Past, present, and future. *Geography Compass*, page e12403, Sep 2018.
- [3] Reginald G. Golledge. Looking back and looking forward. *Geographical Analysis*, 31(4):318–323, Sep 2010.
- [4] Chris Anderson. The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 2006.
- [5] Eric S. Raymond. *The Cathedral and the Bazaar*. O'Reilly, 1999.
- [6] John Mark. Why open source failed. <https://medium.com/@johnmark/why-open-source-failed-6cae5d6a9f6>, July 2018.
- [7] Sergio J. Rey. Show me the code: Spatial analysis and open source. *Journal of Geographical Systems*, 11:191–207, 2009.
- [8] Geolytics. Neighborhood change database 1970-2000. <http://www.geolytics.com/USCensus,Neighborhood-Change-Database-1970-2000,Products.asp>, 2018.
- [9] IPUMS. Ipums. <https://www.ipums.org/>, 2018.
- [10] John R. Logan, Zengwang Xu, and Brian J. Stults. Interpolating U.S. decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. *The Professional Geographer*, 66(3):412–420, May 2014.

- [11] Sergio J. Rey. Open regional science. *Annals of Regional Science*, 52(3):825–837, 2014.
- [12] N. Eghbal. *Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure*. Ford Foundation, 2016.
- [13] May Yuan. Representing complex geographic phenomena in GIS. *Cartography and Geographic Information Science*, 28(2):83–96, 2001.