Commentary: general principles and analytical frameworks in geography and GIScience

Michael F. Goodchild
Department of Geography, University of California, Santa Barbara, CA, USA

Some initial thoughts

Geography and GIScience (geographic information science) are both concerned as disciplines with the infinite complexity of the surface and near-surface of the Earth, or what we might call the geographic domain. Many other disciplines also concern themselves with this domain, including most if not all of the social and environmental sciences, but none do so with the generality of geography and GIScience. Geography has a long tradition of concern with integration, with exploring the links that exist between disciplines and with problems whose solution requires knowledge that extends across many disciplines.

It is not surprising, therefore, that an invitation to address the general principles and analytical frameworks in geography and GIScience has generated such a diversity of perspectives. There are clearly many questions one might ask about the geographic domain, and many routes to building representations that might be used to address those questions, especially when those representations must capture many distinct phenomena in the same framework. Geographers have long used maps as a framework with which to create, store and share representations of the geographic domain. But maps have obvious limitations: they are flat while the geographic domain is curved; they use two spatial dimensions to represent the three spatial dimensions of the domain; they must necessarily focus on static features; unlike numerical data, they are not readily submitted to quantitative analysis; and the scale of a map imposes a constraint on the representation’s level of detail.

Today, the move to digital representations has in principle removed many of these limitations. Geographic information systems (GIS) and spatial databases now capture, represent and analyse the information that was previously shown in maps; they include the third spatial dimension; and it is now possible to represent and investigate time-dependent phenomena. Thus, **tupu**, the concept advanced by Chen Shupeng and the subject of Li’s paper (Li, this volume), is in many ways the guiding principle of today’s spatiotemporal databases. Although there have been very important advances in the capturing of greater detail, spatial and temporal resolution must always remain limited to some degree because of the limitations of our observing systems. Moreover, practice is often slow to adjust to new opportunities, and many of the decisions made in the early days of the digital transition, at a time when computational resources were extremely limited, still have their legacy effects today (Goodchild 2018).

Clearly, any general principles that might apply to the geographic domain would be extremely valuable as a basis for digital representation and analytic frameworks, and several are identified in these papers. Many make reference to the principle of spatial dependence, nicely expressed by Tobler (1970) in what he suggested might qualify as a First Law of Geography: nearby things are more similar than distant things. The practice of mapping topography with contours would be impossible without it, as would the practice of dividing the world into areas of approximately uniform characteristics – the regions of regional geography or the polygons of GIS. Anselin (1989) argued that spatial heterogeneity was also a defining principle, a theme pursued by Fotheringham and Sachdeva (this volume) in their discussion of geographically weighted regression (GWR). Jiang (this volume) argues for scaling as a principle, based on the observation that small geographic phenomena tend to be much more abundant than large ones and that abundance is often almost precisely related to size by a power law. Central to all of these discussions is the concept of geographic context, or the tendency for geographical surroundings to influence outcomes. This is one possible basis for the similarity principle advanced by Zhu and Turner (this volume), and for the spatial heterogeneity discussed by Fotheringham and Sachdeva (this volume). Finally, scale and its related
concepts of spatial resolution, granularity and representative fraction continue to spark novel developments (Phillips, this volume)

**Laws, principles and norms**

As a discipline, geography has long struggled with defining its central purpose: should it be to describe the unique characteristics of places (an idiographic focus), or to find laws or principles that apply everywhere in the geographic domain and at all times (a nomothetic focus)? Is the discipline concerned with description or with uncovering universals (or perhaps with a mix of both)? The so-called quantitative revolution of the late 1950s and 1960s represented something of a high point in this endless debate, when the nomothetic cause appeared to be in the ascendancy for more than a decade and leaders such as Brian Berry, Peter Haggett, David Harvey, William Warnitz, William Bunge, Peter Gould and many others promoted geography as a scientific discipline searching for the general principles that supposedly lie hidden in the complexity of the geographic domain, and fuelled in part by the introduction of computers and the popularization of statistical methods.

A desire for laws, or ‘physics envy’ as it might be put facetiously, can be found in many of the environmental and social sciences (for a discussion of laws in ecology, see Dunn 2021). Yet unlike the laws of physics, given the complexity of the geographic domain any pattern that might appear to qualify as a law will certainly not be deterministic. We expect exceptions to Tobler’s First Law (TFL), and we recognize that what is true at one location may not be true elsewhere, and at other times. In their paper Kedron and Holler (this volume) argue that spatial heterogeneity may lead to a failure to replicate across space, and Goodchild and Li (2021) have suggested a concept of weak replicability when parts of a discovery at one location can be generalized to other locations but other parts of the same discovery cannot. GWR fits this model, as it maintains a single model in the form of a linear equation, but allows the coefficients of the model to vary spatially.

Whether the term ‘law’ is justified in such cases has often been debated (see, for example, the papers in the forum on TFL edited by Sui 2004). ‘Principle’ is a possible alternative, but carries no specific meaning. Instead, I propose the term ‘norm’, implying an ideal or model against which reality can be compared, with a meaning somewhat similar to the ‘stylized facts’ identified by Kaldor (1961) when writing in economics. We refer to many techniques of design as normative, meaning that they seek to construct an improved world according to some ideal principle; for example, location–allocation methods design locations for central facilities according to some principle of optimality, such as maximizing coverage (Church and Murray 2018). Here I suggest that the term might be used more broadly, beyond the context of design, where it would not necessarily imply improvement, but a comparison between reality and some abstract and universal set of assumptions.

Suppose, for example, that we were to see TFL as a norm. Displaying data in map form would immediately allow the eye to identify anomalies – locations where the observed value is not similar to its neighbours, either by being high but surrounded by low values, or being low and surrounded by high values. We might be led to seek explanations for each anomaly, in terms of other properties that are true of its location or of its surroundings. In this way, local context is recognized as producing variations around the norm, in the spirit of spatial heterogeneity. From this perspective, TFL becomes a lens through which to explore spatial data, and an example of a theoretical geography which is neither idiographic nor nomothetic.

The literature of geography is full of such norms. Central place theory (Christaller 1933, 1966), for example, makes certain simple assumptions about the settlements that provide retail and marketing functions to a dispersed population, and from those assumptions reaches a set of conclusions about the sizes and spatial distribution of settlements. As a theory, it must be regarded as a failure since it has failed in tests against actual settlement patterns. But as a norm it allows the researcher to focus on why the observed settlements fail to match the theory’s predictions of hexagonal spacing and a hierarchy of different settlement functions. Does the theory fail because the assumptions of a uniform plain and uniform density of the dispersed population are inaccurate, or because people do not always visit the closest offering of a retail good, or because travel is not equally possible in all directions, or because of some special circumstances at certain locations? In my own neighbourhood of Seattle there are no fewer than three large grocery stores within a few minutes’ walk. Again, what aspect of the theory’s assumptions would have to be violated to explain this?

Scaling laws provide another example, especially the scaling of city populations observed by Zipf (1949) and often known as the rank–size rule: that is, a power–law relationship between a city’s population and its rank. I have analysed a range of geographic phenomena, and find that many follow a more or less precise power law, but there are interesting exceptions. The world’s largest lakes by area follow a power-law closely, but the world’s tallest mountains do not, since there are
many peaks that are almost as high as Mt Everest, and in the conterminous US there are several peaks (Mt Rainier, Pike’s Peak) that are almost as high as Mt Whitney. As with TFL, the nomothetic question of whether phenomena in general follow the norm is not as interesting as the question of why certain phenomena so do not. One simple explanation is that the Earth’s liquid core limits the potential growth of mountains on its crust, but no such constraint applies to lakes. I have also analysed my flights over a ten-year period, and have found that the number of landings at different airports follow a quite precise rank-size rule with one exception: the Rank 1 airport (Santa Barbara) falls well short of what the rule would predict, for the simple reason that almost all trips from Santa Barbara must pass through the Rank 2 airport (LAX) or the Rank 3 airport (SFO).

I like to think of the spatial analyst as armed with a quiver of norms. Some, such as the scatterplot, are widely used in the sciences and not specifically spatial, but they serve to identify anomalies in the form of positive and negative residuals. When applied spatially, they lead to immediate questions: what might explain why that location has a positive residual, or that other location has a negative one? Others, however, are specifically spatial, including TFL, central place theory, the Von Thünen theory of agricultural land use, the distance-decay functions of spatial interaction modelling, Horton’s law of stream number, and the rank–size rule. These are not so much failed theories as norms against which the real world can be compared, and used to connect observations to the theory’s assumptions. In short, I am suggesting that the purpose of the social and environmental sciences, and more specifically of geography and GIScience, should not be the discovery of nomothetic and deterministic laws since such laws are unlikely to exist in the complexity and spatial heterogeneity of the geographic domain, nor the pure description of idiographic study. Instead, much of the power of geographic inference stems from the comparison of reality against a set of well-defined norms that are based on simplifying assumptions.

A concluding comment

Over the past few decades, the discipline of geography has seen many fundamental changes. The geographer’s powers of representation and analysis have advanced very rapidly as a result of the digital transition, the advent of geographic information systems and the rigour introduced by GIScience. These six papers have demonstrated that much remains to be done, or as Jiang argues, to develop the further advances that will ‘ensure a prosperous future for the new geography’. Many would argue that geography and GIScience are already prosperous, because of the important roles that they play in the solution of practical problems, the support of day-to-day decision-making and the generation of new knowledge about the geographic domain. But clearly there is room for even greater prosperity, as these six papers show.

I have argued here that the search for general principles, in emulation of the ‘senior’ sciences of physics and chemistry, will always be frustrating in the enormous complexity of the geographic domain and the abundance of social and physical processes that help to shape it. Yet we have a large library of what I have termed norms – principles if you prefer – that provide a theory-based comparison and allow us to identify local exceptions and departures and to connect these with failures in the specific assumptions that underpin the norms. This is not ‘the exception that proves the rule’ so much as the rule that exposes the exceptions, a lens through which it is possible to identify and interrogate those exceptions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Anselin, L. 1989. “What Is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis.” Technical Report 89-4. Santa Barbara, CA: National Center for Geographic Information and Analysis. https://escholarship.org/uc/item/3ph5k0d4 (accessed 30 December 2021).

Christaller, W. 1933, 1966. “Die Zentralen Orte in Süddeutschland.” In Gustav Fischer Verlag, Jena, Central places in southern Germany. edited by C. W. Baskin, Englewood Cliffs NJ: Prentice Hall.

Church, R. L., and A. Murray. 2018. Location Covering Models: History, Applications and Advancements. Cham, Switzerland: Springer.

Dunn, R. 2021. A Natural History of the Future: What the Laws of Biology Tell Us about the Destiny of the Human Species. New York: Basic Books.

Goodchild, M. F., and W. W. Li. 2021. “Replication across Space and Time Must Be Weak in the Social and Environmental Sciences.” Proceedings of the National Academy of Sciences 118 (35): e2015759118. doi:10.1073/pnas.2015759118.

Goodchild, M. F. 2018. “Reimagining the History of GIS.” Annals of GIS 24 (1): 1–8. doi:10.1080/19475683.2018.1424737.

Kaldor, N. 1961. “Capital Accumulation and Economic Growth.” In The Theory of Capital, edited by F. A. Lutz and D. C. Hague. New York: St Martin’s Press.

Sui, D. Z. 2004. “Tobler’s First Law of Geography: A Big Idea for a Small World?” Annals of the Association of American Geographers 94 (2): 269–277. doi:10.1111/j.1467-8306.2004.09402003.x.

Tobler, W. R. 1970. “A Computer Movie Simulating Urban Growth in the Detroit Region.” Economic Geography 46 (Supplement): 234–240. doi:10.2307/143141.

Zipf, G. 1949. Human Behavior and the Principle of Least Effort. Cambridge: Addison-Wesley.