

METHODS: FIRST LAW OF GEOGRAPHY

Michael F. Goodchild, University of California, Santa Barbara

Michael F. Goodchild  
Department of Geography  
University of California  
Santa Barbara, CA 93106-4060  
USA

# METHODS: FIRST LAW OF GEOGRAPHY

**Keywords:** nomothetic science, idiographic science, positivism, spatial sampling, spatial dependence, spatial autocorrelation, geostatistics, semantic interoperability, spatial heterogeneity

## Glossary

geostatistics: the theory of regionalized variables, or spatial variables subject to spatial dependence

idiographic science: the scientific description of the unique properties of places and times

nomothetic science: science that stresses the value of general principles that apply everywhere in space and time

positivism: a philosophy of science that stresses the confirmation and refutation of hypotheses through experiment

semantic interoperability: successful sharing of data as a result of terms having shared meaning

spatial autocorrelation: a measure of spatial dependence

spatial dependence: the tendency for nearby things to be more similar than distant things

spatial heterogeneity: the tendency for conditions to vary over a space

## Synopsis

Tobler's First Law was a product of the quantitative revolution of the 1960s, and efforts to turn geography into a nomothetic science. It was largely ignored as the quantitative revolution declined, but has recently gained prominence with the growth of GIS. Despite notable exceptions, it is hard to imagine a world in which it is not true, and it provides a very useful principle for the design of geographic information systems and for the spatial analysis that such systems support. The law clearly applies in certain other spaces. Spatial heterogeneity is proposed as a second law. The discipline of geography today seems willing to accommodate a variety of views on the significance of the law.

## Introduction

In science generally, it is widely accepted that the most valuable truths about the natural world are those that apply everywhere in space and time. There would be little value, for example, in a periodic table of the elements that was valid only in North America, or only on Tuesdays. Such knowledge is described as nomothetic, and philosophies of science such as positivism see it as the end product of a lengthy but well recognized process of inference and experimentation. By contrast, knowledge about the specific properties of places and times is described as idiographic, and many of the terms used to characterize such knowledge are pejorative – consider, for example, descriptive, anecdotal, and journalistic.

The notion that there might be nomothetic truths within the domain of the discipline of geography has risen and fallen in popularity over the years. In the 1950s Hartshorne and

Schaeffer championed the idiographic and nomothetic positions respectively, and Schaeffer inspired a generation of graduate students, many of them enrolled at the University of Washington, to direct their research towards nomothetic knowledge, in areas such as central place theory and theoretical geomorphology. Bunge's 1962 book *Theoretical Geography* (and its second edition of 1966) captured this movement most forcefully, arguing that geographic knowledge "must meet certain standards including clarity, simplicity, generality, and accuracy" (p.2) and that the purpose of geographic research was to discover "these patterns, these morphological laws...so that (our) planet, Earth, fills (our) consciousness with its symmetry and ordered beauty" (p.xvi, emphasis added).

While the nomothetic impulse flourished for more than a decade, it fell into decline in the 1970s with the waning of the quantitative revolution. But in the 1990s a degree of revival took place, and much of the earlier work became newly fashionable, motivated at least in part by the rise of geographic information systems (GIS). One of the early observers of this shift was Peter Taylor, who argued in a 1990 critique that GIS represented "the positivists' revenge".

GIS is a generic tool designed to represent a vast array of distinct types of geographic information. As such it must recognize the nature of geographic information, and exploit its significant features in order to achieve both generality and acceptable performance. Thus the nature of geographic information is an important topic in GIS texts, and the growth of interest in GIS has prompted new research into principles, regularities, and laws that can be found through experimental investigation. Of these, by far the most important has turned out to be a principle suggested almost tangentially in a paper written in 1970 by Waldo Tobler, one of the leading members of the University of Washington school, on the topic of modeling the urban growth of Detroit. Tobler suggested somewhat tongue-in-cheek that the principle might be termed the First Law of Geography, but in recent years as its generality and utility has become recognized a consensus has emerged that it does indeed justify that title.

The next section describes the law in detail, and discusses its generality and utility. The following section reviews recent commentaries, and examines the case for additional laws.

### **Tobler's First Law of Geography**

The Earth's surface is almost infinitely complex, and it would be impossible to characterize even a small part of it with perfect fidelity. Instead, geographers and others often indulge in spatial sampling, selecting a comparatively small number of locations at which to collect data, and assuming that the gaps between such observations can somehow be filled. If every location had characteristics that were completely independent of those of its nearby locations, this would clearly be impossible. But in reality characteristics tend to vary fairly slowly over the Earth's surface, such that the characteristics at one location tend to be similar to those at nearby locations. Of course, what exactly is meant by "nearby" and "similar" remains to be seen, and depends on the particular characteristics of interest. Weather, for example, tends to vary very little over

distances less than 1km, but to vary greatly over distances of 1000km; soils, on the other hand, can vary substantially over distances as short as 10m.

Tobler originally stated the First Law (TFL) in the form “All things are related, but nearby things are more related than distant things”. The important message is in the second part, and because “related” carries unnecessary connotations of causality TFL is better stated as “nearby things are more similar than distant things”. TFL thus forms the basis of the process known as spatial interpolation, or the formal process by which the gaps produced by spatial sampling can be filled.

The word law is very powerful, suggesting the kind of universality that only a Second Law of Thermodynamics can possess. Nevertheless there are clearly exceptions even to the Second Law of Thermodynamics, as Maxwell’s Demon demonstrates. Newton’s Laws of Motion were found to have exceptions at both ultra-small and ultra-large scales, but despite quantum mechanics and the Special Law of Relativity it is still common to refer to them as laws. In short, as commonly used in science the term law does not necessarily imply perfect predictability.

In the social sciences the concept of law is much more controversial. Many would argue that law-like behavior will never be found among human populations; and even that search for it is unethical. The experimental physicist Ernest Rutherford is once said to have remarked, perhaps after having been annoyed by some social-scientist colleagues, that the only truth that could possibly be discovered in the social sciences is “some do, and some don’t”<sup>1</sup>.

Against this background it seems fair to label Tobler’s statement as a law, even though it is far from deterministic, and even though its parameters vary from one phenomenon to another. There are clearly exceptions to the law, when phenomena reveal less similarity locally than at a distance. The checkerboard is often cited as an example, since every white square is surrounded by the opposite color, but note that in order for this to occur it is necessary for every square to have a uniform color; in other words TFL applies at the within-square scale, but not at the between-square scale. On the economic landscape the territories established by certain organisms similarly defy TFL, since the presence of one organism makes others less likely in the immediate vicinity. The cliffs of the Grand Canyon represent another exception to TFL, since the property elevation varies dramatically over short distances.

One way to gain insight into the significance of TFL is through a simple thought experiment. In a world in which TFL was absent nearby places would be as different as distant places. It would be necessary to step only a vanishingly small distance away from one’s current location to encounter the full range of conditions on the Earth’s surface, from the height of Mount Everest to the depths of the Marianas Trench, and from the summer temperatures of Death Valley to the winter temperatures of the South Pole. A world without TFL would be an impossible world from the perspective of human existence.

---

<sup>1</sup> While this comment is frequently cited, the author has been unable to locate an original source.

Several methods have been devised for characterizing TFL and the property it addresses, which is known generally as spatial dependence. The field of geostatistics, also known as the theory of regionalized variables, was devised by statisticians as a way of formalizing the principle known informally as TFL, and of optimizing the process of spatial interpolation that relies on it. The variogram (or more correctly semivariogram) is defined as a function describing the increase of variation in a phenomenon with increasing distance. In the absence of TFL variation over short distances is as large as variation over long distances, and the variogram is flat. With phenomena that obey TFL, however, variation is generally observed to rise monotonically until it reaches a maximum value known as the sill, at a distance known as the phenomenon's range (Figure 1).

[Figure 1 about here]

TFL can also be measured at fixed scales for patterns such as the checkerboard, or for data distributed over irregularly shaped regions such as states or counties. Two indices of spatial dependence are in common use: Moran's  $I$  and Geary's  $c$ . Both are described more specifically as metrics of spatial autocorrelation. Moran's  $I$  scales from positive values when TFL applies, to zero when neighboring values are as different as distant values, to negative when neighboring values are more different than distant values. The respective ranges for Geary's  $c$  are less than one, one, and greater than one. Despite the existence of two competing statistics, however, the terms used to describe spatial dependence typically follow the Moran ranges, and patterns for which TFL is valid are commonly described as having positive spatial dependence. Figure 2 shows such a data set and its associated Moran statistic.

[Figure 2 about here]

Spatial dependence can be described, largely correctly, as the most important property of any spatial pattern. It can be useful in distinguishing the impacts of hypothesized processes, since the presence of strong, positive spatial dependence at a particular scale implies that the processes causing the phenomenon are similarly persistent at that scale. Smoothing processes such as glaciation and diffusion processes such as migration both result in patterns with strongly positive spatial dependence, while sharpening processes such as economic competition result in the opposite. But this kind of inference is scarcely sufficient to explain the growth of interest in TFL in recent years, which relies instead on much more utilitarian arguments concerned with GIS design.

In a world without TFL every point's characteristics would bear no relationship to those of its neighbors, and in order to represent any geographic data set in a GIS it would be necessary to characterize every point. The result would be an impossibly large data set. TFL allows spatial patterns to be captured by sampling, since spatial interpolation can always be used to fill in the gaps. TFL is the basis on which all contour maps are made, and the basis on which each day's weather maps are compiled from point data. It allows large areas to be characterized as homogeneous, and represented in a GIS as polygons

rather than as multitudes of points, achieving massive degrees of data compression. In other words, TFL is the basis of the armory of tricks with which GIS databases represent what is in principle an infinitely complex world. The specific details of any phenomenon's spatial behavior, as represented for example in its variogram, are the basis on which the density of sampling is determined, since points spaced closer together than the range of the phenomenon will yield observations that are to some degree redundant. It is also the basis for such parameters of GIS representation as pixel size, minimum mapping unit, and spatial resolution. No wonder, then, that the growth of GIS stimulated a renewed interest in TFL and its implications.

## **Commentary**

In 2004 the *Annals of the Association of American Geographers* published a Forum on TFL, with five distinct perspectives, an introduction by the Forum's editor Daniel Sui, and a response by Waldo Tobler. Jonathan Phillips reinterpreted TFL in the context of physical geography, while Harvey Miller discussed the potential for greater formality in TFL as scientific understanding of social behavior advances. Among the skeptics, Trevor Barnes argued that TFL, as all science, must be understood as a social construction that is intimately related to the places of its origin and the life of its author, while Jonathan Smith examined TFL against three requirements -- universality, necessity, and synthesis -- and found it lacking in all respects. It is clear that TFL is a contentious idea within the discipline, particularly among cultural geographers, and that debates over its meaning will continue.

The present author's contribution to the Forum addressed two questions: does TFL apply to spaces other than the geographic, and are there other laws of geography waiting to be discovered? On examination, it is clear that TFL applies to the spaces of other planets, but interesting insights can be gained by asking whether it applies to the space of the human brain, the human genome, or the digits of pi. Similar benefits to those obtained by exploiting TFL in the representation of geographic information, including compression and interpolation, might apply to branches of informatics that address these other spaces.

Several candidates for additional laws have been suggested. Perhaps the strongest candidate is spatial heterogeneity, the observation that conditions on the Earth's surface vary, in what a statistician would describe as non-stationarity. If it is impossible to conceive of perfect explanation in the social sciences, and indeed in many of the environmental sciences as well, then it follows that the unexplained variation will display geographic patterns, and that models estimated for different parts of the geographic world would have different parameters. This concept has been addressed recently through the development of a range of place-based statistics, or statistics that measure the properties of places, with no expectation that the results will generalize to other places. They include Geographically Weighted Regression, a technique that allows the parameters of a linear model to vary spatially, leading to interesting insights in many applications.

Spatial heterogeneity has an interesting implication for another aspect of GIS, the problem of interoperability. Two systems are said to be interoperable if data from one can be read and processed by the other, and clearly issues of data format are at the forefront

of the problem. More difficult to address, however, are issues of semantics, or the variation in the meaning of terms from one part of the world to another. A number of projects, such as the European Union's INSPIRE, have attempted to overcome such variations through the establishment of common standards, and techniques for translation and cross-walking between classifications. But the problem remains a serious impediment to the integration of geographic data, and to studies that span several jurisdictions.

Lack of semantic interoperability, through the adoption of different classification systems and different meanings of terms in separate jurisdictions, can be readily understood as a necessary consequence of spatial heterogeneity. Left to their own devices, researchers in any limited area of the Earth's surface will adopt systems that work well for that area; and because of spatial heterogeneity, are necessarily different from the systems that work well in other areas. Global standards will never work as well, and there will always be resistance locally to their adoption.

## **Conclusion**

Seen from half a decade later, the discipline of geography of the 1950s appears to have been remarkably homogeneous. The debates of that period led to the divisive quantitative revolution, which was followed by an equally divisive retreat. Today, however, the discipline is thriving, at least in the U.S., as a polyglot of vastly different perspectives and methodologies. The reaction to TFL, as reflected in the *Annals* Forum of 2004, is only one indicator of the degree to which the discipline has agreed to tolerate, or at least agreed to differ. Within the GIS community TFL is regarded as a cornerstone of representation and a key to analytic insights, while large sectors elsewhere in the discipline may raise fundamental objections to the very idea of laws in geography.

Any effective use of GIS requires an understanding of the nature of geographic information, and TFL stands as the most compelling summary of that nature. It demonstrates once again that the discipline of geography is the logical home of GIS, because only with a deep understanding of the nature of the real geographic world can one hope to make sense of the bits and bytes of its digital representation.

## **Further reading**

- Bunge, W. (1966) *Theoretical geography*. Second Edition. Lund: Gleerup.
- Cliff, A. D. and Ord, J. K. (1981) *Spatial processes: models and applications*. London: Pion.
- Fotheringham, A. S., Brunson, C., and Charlton, M. (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Hoboken, NJ: Wiley.
- Isaaks, E. H. and Srivastava, R. M. (1989). *Applied geostatistics*. New York: Oxford University Press.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W. (2005) *Geographic information systems and science*. Second Edition. New York: Wiley.
- 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.

Taylor, P. J. (1990) Editorial comment: GKS. *Political Geography Quarterly* 9(3): 211-212.

Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234-240.

### **Figure captions**

1. A typical variogram analysis of Tobler's First Law. The map on the left shows precipitation amounts at points for California's Bay Area. The graph on the right plots variance against distance, showing the familiar monotonic rise.

2. Percent black mapped for Milwaukee census tracts from the 1990 census. Moran's *I* statistic is 0.8971, indicating very strong positive spatial dependence (and adherence to TFL).