



# Introduction to the special issue on scale and spatial analytics

Michael F. Goodchild<sup>1</sup> · Wenwen Li<sup>2</sup> · Daoqin Tong<sup>2</sup>

Received: 14 July 2022 / Accepted: 15 July 2022 / Published online: 28 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## 1 Background

The surface of the Earth covers some 510 million square kilometers, so a vast amount of digital information would be required to represent it in any detail. But for many purposes it would be sufficient to work from a coarse representation. To create an outline map of the continents, for example, one could omit all detail about the oceans, topography, and human habitation, and capture only a set of lines that follow the coasts (and the land boundaries between Europe and Asia and between North and South America). An additional decision would have to be made about the appropriate level of detail in the representation of coastlines, because the convolutions of many coastlines and their offshore islands might not be relevant to a particular application.

It follows from this basic argument that level of detail is a fundamental property of all geospatial data sets, yet there are many ways of defining and measuring level of detail. Many of these are associated with the term *scale*, but that simple English term is used in many ways, not all of which address level of detail. Specifically, scale is often used to refer to the geographic extent of a study; a study of a small area might be described as small-scale, while a large-scale study would address a large geographic area. However, this special issue is concerned solely with scale's meaning as level of detail.

The importance of scale as a topic in geography and geographic information science is reflected in the number of review books that have appeared over the years (for example, Quattrochi and Goodchild 1997; Sheppard and McMaster 2004; Zhang et al. 2014) and the entries on scale in encyclopedias and manuals, which might suggest that the topic is well understood. Recently, however, scale has taken on renewed significance in spatial analytics, for several reasons. First, as a result of the widespread engagement with social media, the use of wayfinding

---

✉ Michael F. Goodchild  
good@geog.ucsb.edu

<sup>1</sup> University of California, Santa Barbara, USA

<sup>2</sup> Arizona State University, Tempe, USA

apps, the launch of satellites capable of capturing images of the Earth's surface with unprecedented levels of detail, and ready access to novel data sets such as StreetView, the spatial analyst today has access to vast resources of geospatial data at levels of detail that were inconceivable as little as a decade ago. These new data are in turn allowing new questions to be asked and new types of research to be pursued. We now know far more about the spatial behavior of individuals in urban areas; far more about the detailed socioeconomic and demographic structure of neighborhoods; and far more about green spaces and urban horticulture, to cite just three applications. Second, new techniques of spatial analysis have been developed and made available to researchers that exploit scale in novel ways. Deep learning, for example, is capable of examining imagery simultaneously at multiple levels of detail, in contrast to traditional methods of analysis of remotely sensed images that work only with the image's original pixel size. This kind of multiscale analysis can be said to emulate the capabilities of the human eye and brain and has proved successful at detecting features in images, such as roads that are obscured by trees and parked cars, that had previously proven resistant to automation.

Scale is also explored through new techniques such as multiscale geographically weighted regression (MGWR), which attempts to identify the scales of various processes operating in the geographic landscape. Finally, new developments in geospatial representation are addressing the difficulties that have traditionally impeded analysis at global scales. These discrete global grid systems (DGGS) create hierarchical structures based on one of the five Platonic solids, allowing geographic variation at any scale to be represented using spatial elements that are approximately equal in size and shape, almost entirely avoiding the complexities of map projections and the mistakes that are often made when the Earth's curvature is ignored. In effect, DGGSs allow variation across a curved surface to be represented at constant level of detail.

These developments led us to organize a two-day international workshop at Arizona State University in February 2020, bringing together a total of 38 participants to discuss Scale and Spatial Analytics. The workshop was organized under the auspices of SPARC, the Spatial Analysis Research Center, as one of its annual research workshops, with funding from Esri, the ASU School of Geographic Sciences and Urban Planning, and SPARC. Full details of the workshop, including the presentations and summaries of discussion, are available at <https://sgsup.asu.edu/sparc/workshop/Scale>. The workshop participants agreed that the topic was timely and that it should be followed by journal articles that explored the issues in depth. Thus two special issues of journals were planned subsequent to the workshop: one focused on environmental issues, to be published in the *Journal of Landscape Ecology*, and this on social issues.

This special issue of the *Journal of Geographical Systems* includes eight papers. Contributors were first asked to submit abstracts, which were reviewed by the editors. Authors of abstracts that were deemed suitable for this special issue were then asked to provide full papers, each of which was subjected to peer review through the procedures of the journal. The next section of this introductory paper provides a review of the papers and explains the order in which they have been organized

in the journal. The final section provides an overview, identifies gaps that remain in our knowledge and understanding of the topic, and points to future research opportunities.

## 2 Review of the special issue papers

The eight research papers included in this special issue cover a breath of literature review, new methodology, applications, and perspectives of scale. The first paper, by Oshan, Wolf, Sachdeva, Bardin, and Fotheringham, provides a comprehensive review on research related to spatial scale, especially quantitative research that enables multiscale analysis. In this vein, the authors define two types of what they term scale multiplicity. Type I refers to multiple definitions of spatial scale, which could be about observation (data), cartographic design, spatial processes, and geographic extent. Type II refers to the combination of multiple scale definitions included in Type I and focuses on identifying underlying scale effects of spatial processes using spatial data collected at different scales. The analysis of Type II scale multiplicity and methods that can reveal such scale effects is a new angle presented in this review paper. The authors analyzed scale-relevant papers from five representative journals in geography and GIScience and found that studies related to process scale only account for 18% of the total research publications surveyed. From these analyses, the authors call for the development of new methods that can enable formal inference about the scale of spatial processes to improve the interpretability and reproducibility of spatial research (Kedron et al. 2021).

Wang and Wu discuss computational challenges in modeling spatial and temporal dimensions of big data (data captured at fine spatial and temporal scales) in discrete-choice models. The authors found that the Bayesian MCMC (Markov chain Monte Carlo) method is among the most computationally efficient approaches in modeling spatial relationships within big data. In dynamic discrete-choice modeling using very large data sets, also known as the modeling of temporal big data, research has centered on better capturing associations of data across different timeframes to enhance model performance. The authors concluded that besides resolving computational challenges, more theoretical and methodological advances as well as sustainable tool development efforts are needed for future innovation.

Su, Dodge, and Goulias analyzed human movement data to understand how varying temporal scales (data sampling rates) would affect estimation of human mobility factors such as travel speed under different transportation modes (e.g., walking, driving, or riding with public transit) and activity space. It was found that a decrease in data sampling rate (i.e., a coarser temporal scale) will cause an underestimation of point-to-point speed. This will increase the uncertainties in detecting speeding violation or identifying a safe speed for motorists. Determining the optimal sampling rate has become an important task to ensure accurate estimation of arrival time in a Mobility-as-a-Service (MaaS) platform where multiple travel modes are combined. The authors also investigated the impact of temporal scale on the estimation of activity space. The results show that methods such as minimum convex polygon

(MCP) are more robust and less sensitive to changes in the temporal scale of the data than are distance-based measures such as kernel density estimation (KDE).

Following these methodological discussions, this special issue includes three interesting application papers that adopt scale-aware analysis. Hohl, Tang, Casas, Shi, and Delmelle integrate temporal-scale modeling into the KDE to predict clusters of events under dynamic background change (e.g., population growth). The authors leverage this approach to detect the space–time patterns of disease risk. Compared to traditional KDE, it can identify areas that are consistently ranked as being at high risk of disease outbreak over time, so that priorities can be given to such areas to perform interventions and disease control.

Cai, Lam, and Zou present a regression model for prediction of land-loss probability in Coastal Louisiana by incorporating neighborhood scale effects. The neighborhood size of each variable, such as distance to the coastline, land fragmentation, and percent of vacant houses, are determined by semivariogram analysis and the resultant neighborhood factors are included into the modeling process. The statistical results show a significant increase in prediction accuracy when neighborhood factors are used in conjunction with land-loss factors (e.g., elevation or median household income).

Ru, Haile, and Carruthers analyzed the correlation between urbanization and improvement of wellbeing and reduction of child growth failure in Sub-Saharan Africa. To achieve this, the authors collected and reconciled multiple years of geospatial data related to urbanization and surveys of children’s health from multiple sources and at different scales, into a uniform spatial frame to exploit their correlations. The regression results show that urbanization helps to reduce child growth failure—children living in urban areas have a reduced chance of being stunted, wasted, or underweight.

Fotheringham and Sachdeva present a new perspective on local spatial modeling by discussing how spatially varying processes may take a dominant effect in causing the discrepancy in the patterns of data aggregated at different geographical scales, an issue known as a MAUP (Modifiable Areal Unit Problem). The authors claim that while MAUP seems to focus on the scale of *data*, its effect in a global model may well be controlled by the spatial *processes* underlying the data. This study also shows that a relationship identified at the global level may well be different from that found at the local level, as the global model and local model are centered to answer different questions. These conclusions reflect the important property of “spatial heterogeneity” in geographical phenomena (Goodchild 2009); they also coincide with the “weak replicability” of research in social and environmental sciences described by Goodchild and Li (2021).

Last but not least, Yuan and McKee present some new thinking about scale in machine learning. The authors use convolution neural networks (CNNs) as an example to showcase how scale is embedded in multilayer representation learning, a process to extract semantically meaningful features (Li 2020). The authors then link this process with geographical representations of fields, object-fields, field-objects and objects, concepts which are considered compatible with information learned from the shallow to deep layers of a CNN. Finally, the power of CNN is demonstrated through an image analysis application in detecting archeological features with varying sizes, shapes, and structures.

### 3 Looking forward

For decades, scale has been an important topic attracting scholars' attention from different fields. It is not uncommon to observe differences when geographical reality is represented with varying levels of detail. Similarly, discrepancies have often been seen in research findings when data are aggregated using alternative scales. Articles included in this special issue provide important insights into issues related to scale, including representation, methods, computation, and applications.

Compared with years ago when data were mainly collected through surveys or field work, nowadays large amounts of data collected using various sensors and devices, and through different platforms, have revolutionized the information available about the Earth. These data can cover a large spatial extent and provide statistically sound samples with a fine spatial and temporal resolution. The unprecedented spatial–temporal coverage, as well as the richness and granularity of big data, provides new opportunities to further advance the study of scale. As an example, some of these data allow elaborate studies to be conducted at the individual level for a better understanding of human mobility patterns. Meanwhile, advances in computation and algorithms dealing with big data offer new ways to examine processes at different scales.

Geographers have mainly been concerned about spatial scale and less attention has been paid to temporal scale and the associated implications. The temporal scales at which spatial processes operate may differ. As Su, Dodge, and Goulias show, human mobility findings may vary with the temporal scale used to aggregate data. Meanwhile, the interaction between spatial and temporal scales may make spatial processes complicated. As highlighted by Wang and Wu, more research and tools are needed to analyze spatial–temporal phenomena. When it comes to real-world applications, there exist jurisdictional scales, and from the government point of view, issues related to a particular spatial process may cut across jurisdictional scales. Formulating effective policies and forming good alignment between scale of management and scale of spatial processes point to another future research direction.

### References

- Goodchild MF (2009) What problem? Spatial autocorrelation and geographic information science. *Geogr Anal* 41(4):411–417
- Goodchild MF, Li W (2021) Replication across space and time must be weak in the social and environmental sciences. *Proc Natl Acad Sci* 118(35):e2015759118
- Kedron P, Li W, Fotheringham S, Goodchild M (2021) Reproducibility and replicability: opportunities and challenges for geospatial research. *Int J Geogr Inf Sci* 35(3):427–445
- Li W (2020) GeoAI: where machine learning and big data converge in GIScience. *J Spat Inf Sci* 20:71–77
- Quattrochi DA, Goodchild MF (eds) (1997) *Scale in remote sensing and GIS*. CRC, Boca Raton
- Sheppard E, McMaster RB (eds) (2004) *Scale and geographic inquiry*. Blackwell, Malden
- Zhang JX, Atkinson PM, Goodchild MF (2014) *Scale in spatial information and analysis*. CRC, Boca Raton

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.