Journal of Spatial Information Science
Number 21 (2020), pp. 83–87

doi:10.5311/JOSIS.2020.21.726

Invited Article

Local modeling: one size does not fit all

A. Stewart Fotheringham

School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA

Received: November 4, 2020; accepted: November 17, 2020

Abstract: This paper considers what happens when we abandon the concept that models of social processes have global application in favor of a local approach in which context or the influence of 'place' has an important role. A brief history of this local approach to statistical modeling is given, followed by a consideration of its ramifications for understanding societal issues. The piece concludes with future challenges and prospects in this area.

Keywords: local modeling, spatial context, process nonstationarity

1 Introduction

A goal of research, in whatever discipline, is to discover something new about the way our world, and our universe, works. We are typically prompted to ask questions such as "why are things like that?" or "what's causing that to happen?" by observing data on something and noting that the values fluctuate over time or over space or both. Such a process goes back to the dawn of civilization when people became aware of the passage of the sun over the course of a day, the changing position of the sun at midday over the course of a year, and the shifting pattern of stars in the night sky. Today, in the spatial realm, we are more likely to be prompted to ask such questions by looking at maps of data, remotely sensed imagery or graphical representations of human behavior. Whatever the source of our questions, they remain basically the same: "why are some values high and some low?" or "why is something present in this location but absent in that one?" Essentially we are asking questions about processes which we cannot observe that have produced *data* which we can observe. The situation is encapsulated in Figure 1.

We measure associations between data that we observe on the real world and from these associations we infer something about the processes that have produced the data we observe. Inference is generally necessary in spatial analysis because the processes we are

© by the author(s) Licensed under Creative Commons Attribution 3.0 License

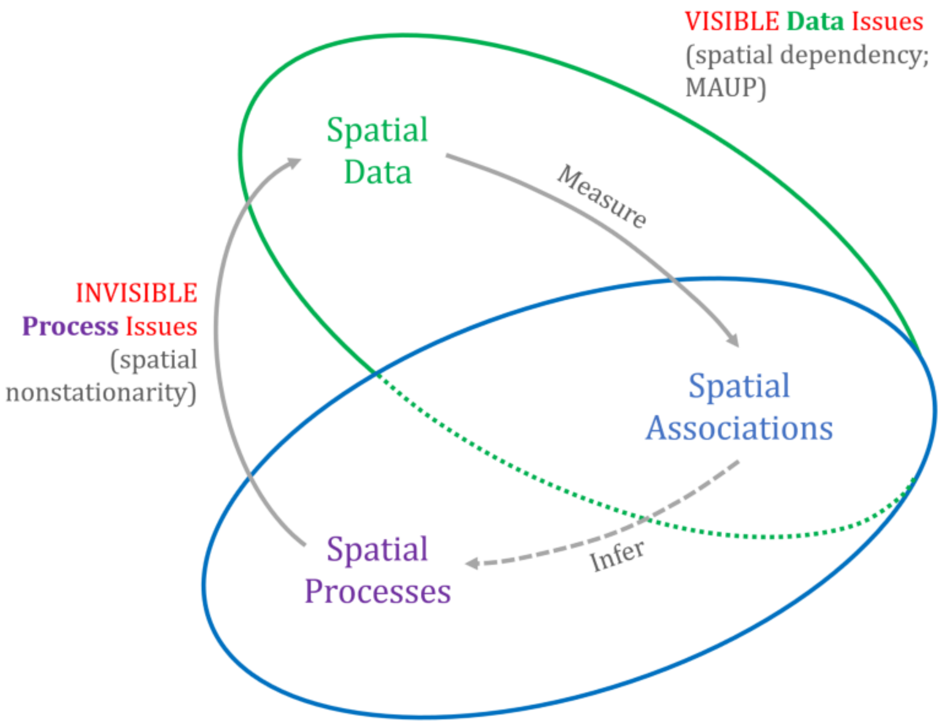

Figure 1: The relationship between spatial data and spatial processes.

interested in are unobservable—we can only see or measure their impacts. Consequently, the subject matter of spatial information science can be divided into the visible and the invisible. Problems specific to spatial data (the visible) include spatial dependency and the modifiable areal unit problem; a problem specific to spatial processes (the invisible) is that of spatial nonstationarity and it is this latter problem that is the focus of the remainder of this paper.

Traditionally, in the analysis of spatial data and spatial processes it was assumed, generally without thought, that the processes that produced the data and associations between data were stationary over space. That is, models were calibrated using data gathered from multiple locations and a single parameter estimate representing each process or conditional relationship was estimated. Consequently, if the processes that were being modeled varied over space, such a single parameter estimate would represent an average of the underlying spatially varying processes, much the same as, say, an average annual rainfall value for a country represents, but hides, the interesting spatial variation in local rainfall values.

The advent of local models in the last couple of decades has altered this perspective by allowing for the possibility that the processes being modeled might vary across space and replacing the single parameters that represented these processes in traditional models with location-specific parameters. Such local modeling frameworks have arisen from different academic and philosophical perspectives but have in common the relaxation of the assumption that processes are stationary over space. Examples of local modeling frame works include Bayesian spatially varying coefficients models [1,5,9], multiscale geographically weighted regression [7,8,13] and eigenvector spatial filtering [10].

2 Implications for research on societal issues

The development of local statistical models has several profound implications for research on critical societal issues. First and foremost is the recognition that, whatever the focus, be it on health-related issues, the study of societal ills, voting preferences, or consumer choice, a *one-size-fits-all* mentality might not be the most appropriate way to understand such issues. In situations involving human behavior where we want to know either if we change x by a given amount, what will the change in y be or what is the best way to achieve a certain change in y, a traditional global approach may be misleading. Where responses to a given stimulus vary over space, more informed decision-making will result from the calibration of models yielding parameter estimates unique to each location.

Secondly, the recognition that processes might vary spatially is a mixed blessing in terms of understanding the transferability of models in the social sciences. On one hand, if we cannot replicate the results of a model calibrated when we use data from a different location, this does not necessarily mean the model is wrong. It could mean the behavior we are modeling is not the same everywhere so the results of the calibration of a model in one location will not be transferable to another location. To aim for models with consistent parameter estimates over space may be a false goal of social science research. On the other hand, if such is the case then models need to be calibrated locally which increases the cost and effort of data collection.

Thirdly, if human preferences and behavior are partly a result of where people live, this raises several questions. To what extent does spatial context affect our decision-making—does it play a major role in our decisions or a minor role and does it vary by the type of decision-making? Over what spatial area does context apply to decision-making—is it at the level of a household, a street, a neighborhood, a city, a region, or a country? How does context affect behavior? Is it local media, the influence of family and friends, local customs, exposures to various societal norms or combinations of all these? Is context simply a shorthand for model misspecification and could we include its effects directly through measurable attributes?

Fourthly, does process spatial nonstationarity explain the modifiable areal unit problem? Most applications of spatial analytical research related to societal problems use aggregated data because individual-level data are generally not available for confidentiality reasons. A well-known problem in dealing with aggregated spatial data is the modifiable areal unit problem (MAUP) which is that the inferences we draw from the analysis of spatially aggregated data can depend on the level to which the data are aggregated [6]. In its most extreme form it is possible to draw completely different inferences from the same underlying data which have been aggregated to different levels. To date, the cause and solution to the MAUP eludes us. However, it is possible to rethink the MAUP as a result of process spatial nonstationarity. If the underlying processes being examined are constant over space, no matter what aggregation of the basic data we undertake, our results should be similar. However, if the processes producing the basic data vary over space, then different aggregations of data will produce different results, and in extreme cases, different inferences about the processes that produced the data. Local modeling may well then alleviate the issue of the MAUP as inferences will not be based on one set of averaged results.

Fifthly, the calibration of local models such as MGWR which yield covariate-specific measures of the scale over which different geographic processes are relatively stable, can inform decision-makers of the geographic extent of areas where certain actions can produce desirable results. That is, rather than having to undertake a global-model-led 'shotgun' approach and applying policies or actions across an entire area, local models can highlight limited subareas where actions should be concentrated to achieve more desirable and efficient outcomes.

3 The future

Local spatial modeling has been part of the spatial analyst?s armory for two decades yet it is still very much in its infancy and evolving rapidly—the advent of MGWR in 2017 being an example [8]. One suspects much more is yet to come. For instance, conceptually one could derive optimised bandwidths as indicators of scale which not only vary across covariates but also vary over space. That is, the area over which a process is relatively stable over space may itself vary over space. Such models would be hugely complex to calibrate and would produce massive amounts of output which would need careful examination for robust interpretation. Inference in local models is a challenge given both the issue of multiple hypothesis testing and the dependency of the tests being conducted [4, 12, 14].

However, the big challenge in local modeling is not statistical but mental. There needs to be a move away from traditional global thinking to an increased recognition that the same action may produce consequences that vary locally. That is, a *one-size-fits-all mentality* is not always the best. This is being recognised statistically with the advent of local forms of analysis other than regression [2, 3, 11] but it has also been recognised more broadly for example in the way that some countries have dealt with the coronavirus pandemic by not imposing blanket restrictions on every part of the country equally but by spatially varying these restrictions. Identifying where limited resources may have the most impact is the key to improving societal well-being and this can only be done through a mentality that recognises that the unobservable processes producing the observable outcomes we want to change are not the same everywhere and need to be examined locally.

References

[1] BANERJEE, S., CARLIN, B. P., AND GELFAND, A. E. *Hierarchical modeling and analysis for spatial data*. CRC Press, 2014.

[2] BOOTS, B., AND OKABE, A. Local statistical spatial analysis: Inventory and prospect. *International Journal of Geographical Information Science 21*, 4 (2007), 355–375.

[3] COMBER, A., BRUNSDON, C., CHARLTON, M., AND HARRIS, P. Geographically weighted correspondence matrices for local error reporting and change analyses: Mapping the spatial distribution of errors and change. *Remote Sensing Letters 8*, 3 (2017), 234–243.

[4] DA SILVA, A. R., AND FOTHERINGHAM, A. S. The multiple testing issue in geographically weighted regression. *Geographical Analysis*, 48 (2015), 233–247.

[5] FINLEY, A. O. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence: Spatially-varying coefficients models. *Methods in Ecology and Evolution 2*, 2 (2011), 143–154.

[6] FOTHERINGHAM, A., AND WONG, D. W.-S. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A 23*, 7 (1991), 1025–1044.

[7] FOTHERINGHAM, A. S., LI, Z., AND WOLF, L. Scale, context and heterogeneity: A spatial analytical perspective on the 2016 US presidential election. *Annals of the American Association of Geographers* (2020).

[8] FOTHERINGHAM, A. S., YANG, W., AND KANG, W. Multi-scale geographically weighted regression. *Annals of the American Association of Geographers 107*, 6 (2017), 1247–1265.

[9] GELFAND, A. E., KIM, H.-J., SIRMANS, C. F., AND BANERJEE, S. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association 98*, 462 (2003), 387–396.

[10] GRIFFITH, D. A. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A 40*, 11 (2008), 2751–2769.

[11] KORDI, M., AND FOTHERINGHAM, A. Spatially weighted interaction models – SWIM. *Annals of the American Association of Geographers 106*, 5 (2016), 990–1012.

[12] LI, Z., FOTHERINGHAM, A. S., OSHAN, T. M., AND WOLF, L. J. Measuring bandwidth uncertainty in multiscale geographically weighted regression using akaike weights. *Annals of the American Association of Geographers 110*, 5 (2020), 1500–1520.

[13] WOLF, L. J., OSHAN, T. M., AND FOTHERINGHAM, A. S. Single and multiscale models of process spatial heterogeneity: Single and multiscale models. *Geographical Analysis 50*, 3 (2018), 223–246.

[14] YU, H., FOTHERINGHAM, A. S., LI, Z., OSHAN, T., KANG, W., AND WOLF, L. J. Inference in multiscale geographically weighted regression. *Geographical Analysis 52*, 1 (2020), 87–106.