

Commentary

Navigating the Methodological Landscape in Spatial Analysis: A Comment on “A Route Map for Successful Applications of Geographically-Weighted Regression”

Taylor M. Oshan

Department of Geographical Sciences, University of Maryland at College Park, College Park, Maryland, USA

The development of “route maps” for spatial analytical methods is a pursuit with important ramifications. Comber et al. propose a route map to guide applications of geographically weighted regression consisting of a three-step primary pathway and a series of secondary arterials. This comment first highlights some concerns about the underlying “map” (i.e., experimental setup and assumptions) and then with the proposed “route” (i.e., core decisions and evaluation criteria). It closes by suggesting a more general focus on identifying modeling issues with the highest impact and facilitating consensus-building, which could improve the future production of route maps for navigating the methodological landscape in spatial analysis.

Introduction

The development of “route maps” for spatial analytical methods is a pursuit with important ramifications. In this case, the “map” is a representation of a scenario an analyst might encounter and the “route” is a series of decisions to be made along the steps of an analytical workflow. Just like a route map for navigating our physical environment, the practical utility of a methodological route map hinges on the assumptions used to create a simplified representation of the real world and the criteria used to determine the efficacy of a route. A judiciously plotted route on an intentionally designed map becomes an invaluable tool for sharing knowledge about how to safely and efficiently get from one location to another, regardless of the type of terrain being traversed. In the methodological world, the result is an outline of best practices that can go a long way toward guiding empirical research and educating future generations of researchers.

Correspondence: Taylor M. Oshan, Department of Geographical Sciences, University of Maryland at College Park, Lefrak Hall, College Park, MD 20740, USA.
e-mail: toshan@umd.edu

Submitted: May 30, 2022. Revised version accepted: August 9, 2022.

doi: 10.1111/gean.12345

179

© 2022 The Author. *Geographical Analysis* published by Wiley Periodicals LLC on behalf of The Ohio State University. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Contributions of this kind are important when they are built on sturdy foundations: those that are consistent, verifiable, and centered on methodological issues with the highest impact.

Comber et al. propose a route map to guide applications of geographically weighted regression (GWR), which is a timely and commendable contribution given the growing popularity of the technique over the past 25 years and the flourishing of local spatial analysis more generally. The proposed “route” consists of a three-step primary pathway “that should always be undertaken” (Comber et al. 2022, p. 1). The initial two steps of the primary pathway are to first consider a traditional linear regression model (LRM) and then second to consider a multiscale variant of GWR (MGWR), both of which are in line with previous recommendations (Oshan, Smith, and Fotheringham 2020). Step three features an additional decision to consider a number of alternative models over MGWR. A set of secondary arterial pathways are also proposed that the authors suggest should be considered only subsequent to the primary three steps, such as collinearity and outliers. This route is sketched across a “mapped” out demonstrative application of modeling soil composition in four analytical scenarios that are created by withholding one or more variables from the dataset. This empirical example illustrates that different models are selected by following the route for each scenario.

It is not clear yet whether the route proposed by Comber et al. or any other single route will be adequate for all GWR applications. Therefore, though most emerging pathways are likely to have some utility, they are also likely to have some segments that can be streamlined or may need course correction. In light of this, this comment proceeds first with some concerns about the underlying map and then with some concerns about the proposed route. It closes with some general suggestions that could improve the production of future route maps for navigating the methodological landscape in spatial analysis.

Some considerations about the map

Cartography is concerned with constructing maps that are accurate and interpretable where accuracy focuses on preserving measurable attributes (i.e., distances, angles, areas) and interpretability focuses on making information on the map accessible to a particular audience. Since every map is a simplified representation of reality, there will always be a trade-off between accuracy and interpretability and this notion also applies to methodological maps. Here, the “map” consists of the experimental setup and associated assumptions used to represent a scenario an analyst might encounter and the trade-off equates to constructing a scenario that is simple enough to generate evidence supporting concise recommendations while remaining realistic enough so that the recommendations are broadly applicable.

It is challenging to define at what point practical guidance can be achieved without considering the nuances that are abundant in the real world. This contention is front and center in the GWR route map (Comber et al. 2022) as the presented application or underlying “map” *is intended* to “provide “real world” practical guidance” (p. 4) rather than advance theory, but *is not intended* “to conduct a nuanced regression analysis” (p. 6). Furthermore, the goal of the application stated within is “to treat each scenario as a distinct and independent data set and not as a linked model specification exercise” (Comber et al. 2022, p. 6) – the provided example is essentially an elaborate metaphor, but one that can easily be mistaken as a model specification exercise. It is important to note then that in practice the analyst will not usually be able to treat subsets of data independently because variable selection is often intertwined with choosing a model specification. This is demonstrated succinctly by the route map itself where the main controlled dimension – the degree of omitted variable bias – is shown to influence the

selection of a model specification. So, the selected map does highlight the influence of omitted variable bias, but it is not clear if this was the primary goal or if it should be. And because an analyst can and should strive to build a model that minimizes omitted variable bias, the induced scenario will not likely hold in the wild and this limits the practical contribution of the example. While the map can be used purely to illustrate the proposed route, it is not clear how it might support the generalization of the route to other practical contexts and modeling issues. It would be helpful to understand if there are scenarios where the route might falter and whether the logic of the route still applies when sketched across a slightly different map. Some stronger evidence therefore seems necessary in order to verify that the route map takes users where they intend to go.

Another feature of the underlying map is that a linear route is assumed to serve as a practical approximation to the sometimes non-linear decision-making process faced by analysts. As a result, a tension arises as Comber et al. suggest both that secondary issues can only be considered after primary model selection and that in practice model building is an iterative process. Unfortunately, both suggestions cannot be adopted simultaneously. Meanwhile, some of the secondary issues could be responsible for generating results with larger practical differences than some of the competing models considered in the primary analysis. This is corroborated by the authors when they state that “secondary analysis will often give rise to different (optimized) bandwidths or a change in the behavior of the bandwidth function to that found with the primary analysis, and thus, potentially changing the chosen GWR form” (Comber et al. 2022, p. 17). This indicates that it may be more important to prioritize what goes into a model than differentiating between similar model forms and it seems conceivable that some of the secondary considerations may be more appropriate as primary considerations or that the linear route map may need to be reformulated as a nonlinear route map.

Some considerations about the route

The task of routing is concerned with identifying pathways on a map that satisfy a set of criteria and the goal is typically to select an optimal route compared to other candidate routes. At the core of routing then, whether using a cartographic map or a methodological map, is the clear definition of these criteria and what is meant by optimal. Here, the “route” consists of decisions under consideration by the analyst and the objective metrics used to inform those decisions. Some additional contentions arise in the GWR route map in this context.

Comber et al. (2022) state that “Critical to Step 3 [of the route] is the presentation and interpretation of the estimated coefficients and associated uncertainties from competing models” (p. 10) but they then proceed by using “only rudimentary assessments of statistical (relationship) significance” (p. 8) even though there exists (M)GWR-specific corrections for assessing significance of relationships (da Silva and Fotheringham 2016; Yu et al. 2020), both of which are also published in this journal and have already been recommended elsewhere as best practices for applied modeling (Oshan, Smith, and Fotheringham 2020). In addition, the route proposes that competing models be compared to each other based on the number of significant coefficient estimates, but they could alternatively be compared by examining whether there is overlap between the confidence intervals for the estimates from each model, which allows uncertainty to be dually considered. Although one model may have a higher number of significant estimates, it could be that the intervals for each estimate overlap and are not effectively different from each other. A recent approach that has become available similarly allows bandwidth estimate intervals to be compared (Li et al. 2020). This strategy could therefore

help to understand whether or not the observed differences between models are significant enough to warrant concern.

Furthermore, the authors posit that the considered models have increased complexity or “flexibility in the specification of the spatial relationships” (p. 3) moving from a LRM to GWR, then mixed (or semiparametric) GWR and finally MGWR (Comber et al. 2022). They also ask “can a simpler model in an LRM, standard GWR, or MX-GWR provide a viable and pragmatic alternative? Or is MS-GWR the only viable option?” (Comber et al. 2022, p. 6). To address this question, it is necessary to have objective metrics to use for comparison, but the terms complexity and flexibility are not subsequently operationalized. They are also perhaps confusing because in the statistical learning literature, which the GWR methodology inherits from, they take on a different meaning with more flexible models being those that can more easily fit to the data and are typically considered more complex because they consume more degrees of freedom (DoF) (James et al. 2013). For example, with 10 covariates, a penalized lasso regression is typically less flexible ($\text{DoF} < 10$) than a traditional ordinary least squares LRM ($\text{DoF} = 10$), which are both generally less flexible than a spline regression ($\text{DoF} > 10$). Using this logic, a broad ordering based on least-to-most flexible would start with the basic LRM, followed by MGWR, then mixed GWR, and finally standard GWR and this could be corroborated by examining the effective number of parameters (ENP) in each model using the same input datasets. Standard GWR and mixed GWR will likely consume excess degrees of freedom and have a higher ENP than MGWR any time there are some spatially varying relationships present that are not all very local. When this condition does not hold it is likely that MGWR would be comparable to its standard GWR and mixed GWR counterparts because they are special cases of MGWR. A question that becomes salient then and seems worth investigating is, “What are the practical tradeoffs to the analyst if the primary route is terminated after the second step”? Toward answering this question, it would be useful to understand if two models, say MGWR and mixed GWR, have slightly different levels of complexity in practice or if one is typically much more complex than the other. Importantly, attributes such as the ENP can be objectively compared using simulations and various empirical applications to test this alternative understanding and accrue evidence over time.

Some suggestions for future route mapping

There are some more general considerations that could facilitate the future production of route maps to navigate the methodological landscape in spatial analysis. First, it would be advantageous to consult the corpus of existing literature through the use of systematic reviews prior to recommending a route in order to identify the pitfalls most frequently encountered in practice. This is the approach used by Oshan, Smith, and Fotheringham (2020) to highlight the areas that could most easily be improved when applying (M)GWR to target the spatial context of the determinants of health outcomes. Second, there are risks associated with making strong recommendations in haste¹; the possibility arises to inadvertently lead analysts astray and obfuscate some important factors that need to be considered in order to accumulate knowledge. Thus, it may be more appropriate to mark the map with individual signposts that help analysts avoid dangerous terrain, rather than suggest specific routes, at least until a consensus emerges towards a prevailing route. Third, there are additional open science practices that can be adopted in order to facilitate consensus-building. For example, the pre-registration of specific hypotheses associated with a proposed route before designing experiments and collecting data to validate

them, as well as an open peer review process. These types of mechanisms can help promote community-driven standards for common analytical goals and systematize best practices.

Note

- 1 For examples in the context of GWR, see Wheeler and Tiefelsdorf (2005), Fotheringham and Oshan (2016) or Lu et al. (2017, 2018., 2019), Oshan et al. (2019).

References

- Comber, A., C. Brunsdon, M. Charlton, G. Dong, R. Harris, B. Lu, Y. Lü, et al. (2022). "A Route Map for Successful Applications of Geographically Weighted Regression." *Geographical Analysis* n/a, n/a. <https://doi.org/10.1111/gean.12316>
- Fotheringham, A. S., and T. M. Oshan. (2016). "Geographically Weighted Regression and Multicollinearity: Dispelling the Myth." *Journal of Geographical Systems* 18(4), 303–29.
- James, G., D. Witten, T. Hastie, and R. Tibshirani. (2013). *An Introduction to Statistical Learning: With Applications in R*, 1st ed.. (Corr. 7th printing 2017 edition). New York: Springer.
- Li, Z., A. S. Fotheringham, T. M. Oshan, and L. J. Wolf. (2020). "Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights." *Annals of the American Association of Geographers*, 110(5), 1–21.
- Lu, B., C. Brunsdon, M. Charlton, and P. Harris. (2017). "Geographically Weighted Regression with Parameter-Specific Distance Metrics." *International Journal of Geographical Information Science* 31(5), 982–98.
- Lu, B., W. Yang, Y. Ge, and P. Harris. (2018). "Improvements to the Calibration of a Geographically Weighted Regression with Parameter-Specific Distance Metrics and Bandwidths." *Computers, Environment and Urban Systems* 71, 41–57.
- Lu, B., C. Brunsdon, M. Charlton, and P. Harris. (2019). "A Response to 'A Comment on Geographically Weighted Regression with Parameter-Specific Distance Metrics.'" *International Journal of Geographical Information Science* 33(7), 1300–12.
- Oshan, T., L. J. Wolf, A. S. Fotheringham, W. Kang, Z. Li, and H. Yu. (2019). "A Comment on Geographically Weighted Regression with Parameter-Specific Distance Metrics." *International Journal of Geographical Information Science* 33(7), 1289–99.
- Oshan, T. M., J. P. Smith, and A. S. Fotheringham. (2020). "Targeting the Spatial Context of Obesity Determinants Via Multiscale Geographically Weighted Regression." *International Journal of Health Geographics* 19(1), 11.
- da Silva, A. R., and A. S. Fotheringham. (2016). "The Multiple Testing Issue in Geographically Weighted Regression." *Geographical Analysis* 48(3), 233–47.
- Wheeler, D., and M. Tiefelsdorf. (2005). "Multicollinearity and Correlation among Local Regression Coefficients in Geographically Weighted Regression." *Journal of Geographical Systems* 7(2), 161–87.
- Yu, H., A. S. Fotheringham, Z. Li, T. Oshan, W. Kang, and L. J. Wolf. (2020). "Inference in Multiscale Geographically Weighted Regression." *Geographical Analysis* 52(1), 87–106.