geographical analysis

Geographical Analysis (2023) **55**, 198–202

Commentary

A Rejoinder to the Commentaries on "A Route Map for Successful Applications of Geographically Weighted Regression" by Comber et al. (2022)

Alexis Comber[1], Paul Harris[2], and Chris Brunsdon[3]

[1]School of Geography, Leeds University, Leeds, UK, [2]Sustainable Agriculture Sciences, Rothamsted Research, North Wyke, Okehampton, UK, [3]National Centre for Geocomputation, NCG, Maynooth University, Maynooth, Ireland

Preamble: The route map

What the RM says

In brief, the GWR Route Map (RM) by Comber et al. (2022a) argues that an ordinary least squares (OLSs) regression and a multiscale GWR should always be undertaken initially. Then, by examining the outputs and results of these, the final choice of model can be determined by applying some very broad rubrics.

Why we wrote the RM

We wrote the RM for two main reasons. First, we have witnessed the increased use of GWR and related approaches such as different types of geographically weighted (GW) regression and other models under a GW framework. This has been documented in, for example, Comber et al. (2022b). We have also been concerned by much recent research, often in domains other than Spatial Analysis and GeoComputation, that have applied only a standard GWR, probably using the ESRI implementation, when thinking on this topic has moved on. As we say in the RM, multiscale GWR (MGWR) should be the default GWR.

What we tried to do in the RM

We sought to provide a guide through the different situations when different flavors of global (fixed coefficient) to local (varying coefficient) regression may be applicable. To support this we decided, after much debate, to use a single dataset and to create a series of discrete studies, referred to as Analysts in the RM, as a conceit to illustrate different situations. These were used to illustrate a process that sought to describe how to decide whether to use a fixed coefficient regression (e.g., a linear model via OLS or restricted maximum likelihood [REML] estimation for an autocorrelated error term) or a varying coefficient regression (e.g., a standard or multiscale

Correspondence: Alexis Comber, School of Geography, Leeds University, Woodhouse Lane, Leeds LS2 9JT, UK.
e-mail: a.comber@leeds.ac.uk

Submitted: October 7, 2022. Revised version accepted: October 18, 2022.

doi: 10.1111/gean.12352

GWR), similar to decisions that may be confronted by the inexperienced or novice user. In these, we tried to outline the GWR considerations to be evaluated by the user in their choice of model.

What we did not address in detail but alluded to

In earlier drafts (and there were many), such as Comber et al. (2020), we tried to describe the full set of *data* and *model* considerations – so called "secondary considerations" – related to investigations, characteristics, and properties of the study data and GWR considerations. These included Exploratory data analysis, Predictor variable specification, Use of spatial predictor variables, Effects of spatial pattern and autocorrelation amongst response and predictor variables, Effects of data preprocessing, Effects of sample size and sample configuration. Similarly, in these earlier drafts we also sought to include full descriptions of specific considerations for GWR: Robustness and outliers, Heteroskedasticity, Autocorrelation, Collinearity, Varying scales of process, Influence of geographical weights, Inference options, and Use as a spatial predictor. For reasons of expediency (paper length, narrative flow) we were unable to detail these in the final submission, although these were alluded to and the reader was directed to the preprint version of the RM (Comber et al., 2020).

Rejoinders to the RM commentaries

We are delighted that the RM paper has stimulated three coherent but diverse Commentaries from leading thinkers in this field (Fotheringham, 2022; Oshan, 2022; Wolf, 2022). Each of these contains robust critiques of the proposed RM and suggest alternative but diverse sets of considerations. We consider each of these in turn and provide a rejoinder by way of response.

Commentary by Wolf

The "The Right to Rule by Thumb" commentary by Levi Wolf (2022) argues that RMs are a controversial practice as they codify or formalize analytical decisions, for which there may be limited consensus amongst the community of practice. It contrasts the arguments used to justify analytical decisions in the GWR RM (the proposed paths through the RM) with *effective* RMs for simultaneous auto-regressive (SAR) and multilevel models, which are based on mathematical and theoretical considerations supported by empirical illustrations and by controlled simulations, spanning decades of research. The author finds the RM problematic because it is empirically rather than theoretically grounded and that as a result that its "situational evidence is simply not precise enough to navigate these forks in the roads." The author argues that a GWR RM should be under pinned by rules generated from examination of the mathematical properties of spatial data and that it should be calibrated through examination of other datasets including simulations, in order to more thoroughly evidence the RM.

Rejoinder to Wolf

This informed commentary provides a framework for a future GWR RM upgrade. It argues that a theoretical and mathematical grounding would vastly strengthen the RM. As described above, it was our intention to provide a guide through the different flavors of GWR and to help decide which of these to use or not. None of the "forks" in the route were intended to be absolute with hard and fast rules (*if this, then that*), rather they were intentionally discursive (Wolf uses the word "fuzzy") with the ambition of making users think about their GWR analysis and not simply plugging and playing – of which evidence abounds in the literature.

Commentary by Oshan

The "Navigating the Methodological Landscape in Spatial Analysis" commentary by Taylor Oshan (Oshan, 2022) indicates that some of the RM segments may need to be streamlined or corrected. It suggests that there is a need for wider debate and consensus building among researchers in this area, to do this, and indicates how the debate could actually be initiated. It also unpicks the "map" and the "route." The map is unpicked with respect to the case study conceits (the Analysts) and the bias associated with omitted variables and their influence on the selection of a GWR model or variant. The author proposes that some examples be included in which the RM might falter, thereby providing stronger evidence to verify the RM. They also suggest the need for iterative consideration of the secondary issues in the model choice. Oshan comments on the route and the interpretation of model outputs (coefficient estimates and their significance) and suggests alternative approaches for this based on examining overlap in confidence intervals in different GWR variants. They propose a different analysis sequence from least-to-most flexible (OLS regression, multiscale GWR, mixed GWR, standard GWR) and identify the need for objective metrics to drive model choice, which could be evaluated and corroborated by examining the effective number of parameters in each model.

Rejoinder to Oshan

This commentary provides suggestions for consensus building, for an iterative approach to GWR model choice and objective metrics as valid ways to tackle this activity moving forward. This and the commentary by Wolf indicate the importance of simulation experiments in guiding the RM. What would be useful is some consensus on the design(s) of the Monte Carlo experiment themselves, so that GWR, its variants and future advances could be objectively compared through time. As for empirical-based studies, there is an analogous RM for simulation-based studies, as simulation models require their own, often difficult, model and parameter decisions also – decisions that can directly impact the results and assessments (e.g., see the appendices in Harris (2019)). Here community development of a simulation toolbox would be useful (see also Comber et al. (2020) for a list of GWR studies using simulation).

Commentary by Fotheringham

The "Alternative Expressway to Defensible Regression-Based Local Modeling" commentary by A Stewart Fotheringham (2022) found the RM to be flawed and provides an alternative 5-point strategy. This contains many contentious statements (the detail of which are not addressed in this rejoinder), but it is substantively argued that (i) statistical models should be robust with regard to their functional form and specification, (ii) that all relevant variables should be included in any model, (iii) that there should be some a priori expectation of and rationale for "some property of a location that could affect any of the relationships being modeled", (iv) that any model that did not satisfy these criteria should not be considered for the local case with a GWR analysis, and (v) that "the only model that needs to be calibrated is MGWR."

Rejoinder to Fotheringham

This commentary seems out of touch with the current data context for much spatial analysis. It is predicated on an assumption that the data we use are bespoke, capture all of the factors associated with the process of interest, and subject to the norms of experimental design. However, this is rarely the case in many real-world applications, including those using the many new forms of data that are usually not collected under formal experimental design. The data we use rarely contain the full set of measurements associated with the process being considered. Instead, data are secondary, messy, incomplete, pulled together from different sources, with varying spatial and sampling properties. This is the new data normal. The commentary also ignores one of the primary roles of the family of GWR models: that of supporting spatial detective work. GWR is an inherently exploratory approach and provides a mechanism to quantifying how and where regression relationships may vary (i.e., via mapping the intercept and the predictor coefficient estimates), including scales of variation through multiscale GWR bandwidths. These can provide spatial insight into the process being considered, can direct further investigation for example by domain experts, or indeed the need for further data collection. Bluntly, we disagree.

Review and prospect

We thank the authors of these commentaries for their efforts, and for taking the time to consider our article in detail. In general, we are pleased to see these – part of our motivation here was to initiate discussion on approaches to modeling spatial non-stationarity in regression models. By setting out one way to proceed through our RM, we intended to make an opening move. One thing we observe from these responses is that there is perhaps a spectrum for motivation for using these kind of models – at one end, an approach that is strongly motivated by underlying theories, and at the other, a more exploratory approach. One also has to consider the idea of data analysis as compromise – the reality of modern data collection is frequently that of "big data" where datasets are large, but quality and suitability assurance are not to the standards achieved by carefully designed surveys or experiments. In many cases, geographical fluctuations in models may be a consequence of this, and spatially varying coefficient methods may act as "spatial detectives" by shedding light on spatial inconsistencies and biases in the data collection, rather than direct measurements of a true underlying process. This suggests the need for a kind of "deep inference" where processes under investigation *and* the process of data collection are considered in equal measure, requiring consideration of underlying process theories, in addition to issues relating to the act of data exploration – perhaps suggesting that the spectrum referred to earlier is something to be scanned, rather than choosing a specific viewpoint from which to carry out analysis.

As we stated earlier, the approach outlined in the GWR RM by Comber et al. (2022a) is not intended to be a strict set of immutable rules, but more of an exemplar of what could be done to respond to a specific research context, and acknowledging that a degree of 'fuzziness' in modeling strategies is inevitable. The replies to our article have been useful in considering potential alternative research contexts, and how they may interact with this kind of fuzziness. We look forward to the debate advancing.

References

Comber, A., C. Brunsdon, M. Charlton, G. Dong, R. Harris, B. Lu, Y. Lu, D. Murakami, T. Nakaya, Y. Wang, and P. Harris. (2020). "The GWR Route Map: A Guide to the Informed Application of Geographically Weighted Regression." *arXiv*: 2004.06070. https://arxiv.org/abs/2004.06070

Comber, A., C. Brunsdon, M. Charlton, G. Dong, R. Harris, Y. Lü, D. Murakami, T. Nakaya, Y. Wang, and P. Harris. (2022a). "A Route Map for Successful Applications of Geographically Weighted Regression." *Geographical Analysis*. https://doi.org/10.1111/gean.12316

Comber, A., M. Callaghan, P. Harris, B. Lu, N. Malleson, and C. Brunsdon. (2022b). "gwverse: A Template for a New Generic Geographically Weighted R Package." *Geographical Analysis* 54(3), 685–709. https://doi.org/10.1111/gean.12337

Fotheringham, A. S. (2022). “A Comment on “A Route Map for Successful Applications of Geographically-Weighted Regression”: The Alternative Expressway to Defensible Regression-Based Local Modeling.” *Geographical Analysis*. https://doi.org/10.1111/gean.12347

Harris, P. (2019). “A Simulation Study on Specifying a Regression Model for Spatial Data: Choosing between Heterogeneity and Autocorrelation Effects.” *Geographical Analysis* 51, 151–181. https://doi.org/10.1111/gean.12163

Oshan, T. M. (2022). “Navigating the Methodological Landscape in Spatial Analysis: A Comment on “A Route Map for Successful Applications of Geographically-Weighted Regression”.” *Geographical Analysis*. https://doi.org/10.1111/gean.12346

Wolf, L. J. (2022). “The Right to Rule by Thumb: A Comment on Epistemology in “A Route Map for Successful Applications of Geographically-Weighted Regression”.” *Geographical Analysis*. https://doi.org/10.1111/gean.12345