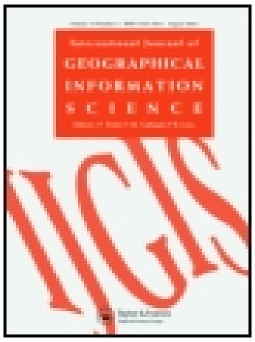

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

A comment on geographically weighted regression with parameter-specific distance metrics

Taylor Oshan, Levi John Wolf, A. Stewart Fotheringham, Wei Kang, Ziqi Li & Hanchen Yu

To cite this article: Taylor Oshan, Levi John Wolf, A. Stewart Fotheringham, Wei Kang, Ziqi Li & Hanchen Yu (2019): A comment on geographically weighted regression with parameter-specific distance metrics, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2019.1572895

To link to this article: https://doi.org/10.1080/13658816.2019.1572895

Published online: 12 Feb 2019.

Submit your article to this journal

Article views: 138

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tgis20

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2019.1572895

Taylor & Francis
Taylor & Francis Group

SHORT COMMUNICATION

Check for updates

A comment on geographically weighted regression with parameter-specific distance metrics

Taylor Oshan[a], Levi John Wolf [b], A. Stewart Fotheringham[c], Wei Kang[d], Ziqi Li[c] and Hanchen Yu[e]

[a]Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, College Park, MD, USA; [b]School of Geographical Sciences, University of Bristol, Bristol, UK; [c]Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, USA; [d]Center for Geospatial Sciences, School of Public Policy, University of California, Riverside, CA, USA; [e]School of Government, Peking University, Beijing, PRC

ABSTRACT

A recent paper in this journal proposed a form of geographically weighted regression (GWR) that is termed parameter-specific distance metric geographically weighted regression (PSDM GWR). The central focus of the PSDM generalization of the GWR framework is that it allows the kernel function that weights nearby data to be specified with a distinct distance metric. As with the recent paper on Multiscale GWR (MGWR), the PSDM framework presents a form of GWR that also allows for parameter-specific bandwidths to be computed. As a result, a secondary focus of the PSDM GWR framework is to reduce the computational overhead associated with searching a massive parameter space to find a set of optimal parameter-specific bandwidths and parameter-specific distance metrics. In this comment, we discuss several concerns with the PSDM GWR framework in terms of model interpretability, complexity, and computational efficiency. We also recommend some best practices when using these models, suggest how to more holistically assess model variations, and set out an agenda to constructively focus future research endeavors.

ARTICLE HISTORY
Received 24 May 2018
Accepted 17 January 2019

KEYWORDS
Multiscale; geographically weighted regression; big models; spatial analysis; spatial statistics

1. Introduction

Lu *et al*. (2016, 2017) develop in this journal a form of geographically weighted regression (GWR) that they term parameter-specific distance metric geographically weighted regression (PSDM GWR), which is then further developed elsewhere (Lu *et al*. 2018). The central focus of the PSDM generalization of the GWR framework is that it allows the kernel function that weights nearby data to be specified with a distinct distance metric, such as Euclidean distance (ED), network distance (ND), travel time (TT) or a parameterized Minkowski distance for each variable in a model. As with the recent paper on Multiscale GWR (MGWR) (Fotheringham *et al*. 2017), the PSDM framework presents a form of GWR that also allows for parameter-specific bandwidths to be computed. As a result, a secondary focus of the PSDM GWR framework is to describe

CONTACT A. Stewart Fotheringham sfotheri@asu.edu
This article has been republished with minor changes. These changes do not impact the academic content of the article.
© 2019 Informa UK Limited, trading as Taylor & Francis Group

a method to reduce the computational overhead associated with searching a massive parameter space to find a set of optimal parameter-specific bandwidths and parameter-specific distance metrics (Lu *et al*. 2017, 2018).

We have two general concerns and several technical concerns with this framework and the body of work it has evolved from. First, the importance of employing parameter-specific distance metrics is exaggerated, which has the consequence of clouding the importance of correctly obtaining parameter-specific bandwidths. Parameter-specific distance metrics appear to provide relatively modest gains in model fit in particular scenarios while potentially diminishing the overall model interpretability. In contrast, parameter-specific bandwidths are the core of MGWR, which generally provides both an increase in model fit and a greater understanding of the processes that generate the observed dependent variable in the model. Second, focusing on the simpler MGWR specification is a more straightforward strategy to achieve similar ends as the PDSM GWR specification. MGWR has lower computational overhead relative to the PSDM GWR specification, increases the interpretability of the model, and comports with other multiscale model specifications (Wolf *et al*. 2017). To further formulate these general concerns, we provide a brief background on the development of the PSDM GWR framework and associated concepts and then expand our reasoning based on several conceptual and technical issues, providing some empirical evidence to demonstrate our concerns and the importance of these issues.

While some evidence is reproduced or extracted from the literature, additional experiments are reported here using the London house price (LHP) data utilized by Lu *et al*. (2014a, 2016, 2017, 2018), as well as the Prenzlauer Berg AirBnB (PBA) rental price data introduced and described in Oshan *et al*. (2018). It should be noted that although the LHP data are made available as part of the *GWmodel* software package (Gollini et al. 2015), it is only a subset (n = 315) of the original dataset used in the literature (n = 2002). Therefore, though the model specifications utilized are the same, some of the results discussed here are necessarily different from those previously presented. Furthermore, while the *GWmodel*[1] package provides some of the necessary functionality to investigate the issues described here, certain procedures are only presently available in the *mgwr*[2] software package (Oshan *et al*. 2018). Wherever possible we indicate which software was used in each experiment, but otherwise strive to keep the focus on substantive issues.

2. Background

The initial basis for the PSDM GWR framework arises from a study that investigated the use of non-Euclidian distance (non-ED) metrics in a traditional geographically weighted regression framework (i.e., not MGWR or PSDM GWR) where only a single distance metric and bandwidth are specified for a given model (Lu *et al*. 2014a). The rationale for exploring non-ED metrics is that different representations of space require commensurate measures of distance to accurately capture spatial processes. Though the substantive differences between three GWR models using ED, ND, and TT, respectively, are very small, the discrepancies between models using ED and non-ED apparently are driven by the spatial distribution of the observations and the shape of the underlying road network (Lu *et al*. 2014a). Therefore, it may be beneficial to employ a non-ED measure in GWR when it is theoretically appropriate for the relationships included in the model. If nearby sites are related to one another over a network, then network or travel metrics make more sense than raw euclidean distances. This is useful when conceptually relevant, despite the fact that the practical impacts of using non-ED metrics are often small (Phibbs 1995, Jones *et al*. 2010, Carling *et al*. 2012). However, in the case that there is no obvious non-ED metric, it does not seem that using an ED metric would be conceptually inapt. Further, unless the data have a highly uneven spatial distribution dispersed across a highly irregular study area, the literature suggests that the numerical results will be substantially similar.

A subsequent study proposed the use of a generalized measure of distance, known as the Minkowski distance (often called l_p *metrics*), that contains Manhattan distance (l_1), ED (l_2), Chebyshev distance (l_∞), and a continuum of other metrics based on the selection of two parameters, p and theta (Lu *et al*. 2016). This data-driven approach optimizes model fit using a GWR-specific Akaike Information Criterion (AICc) to simultaneously select the two Minkowski parameters, along with the local parameter estimates and an optimized bandwidth, but has the effect of significantly increasing the computational overhead of the model calibration. The methodology is evaluated on three simulation scenarios with known coefficient surfaces that are: (i) spatially smooth across a regular grid; (ii) anisotropic and irregularly spatially distributed; and (iii) randomly distributed across a regular grid. Though the Minkowski GWR approach is shown to achieve a marginally lower AICc value and sum of squared error between the known and estimated parameters, there are some flaws in the experimental design that are discussed below.

Coincidental to the use of non-ED metrics in GWR, Yang (2014) in her PhD thesis proposed flexible-bandwidth (FB)GWR which allows parameter-specific bandwidths to be estimated and which is a precursor to the more fully developed Multiscale (M)GWR (Fotheringham *et al*. 2017). These latter two models are focused explicitly on providing an indicator of scale (bandwidth) for each process surface which can enhance intuition regarding the spatial relationships in the model and provide improved model fit. The concept of allowing aspects of spatial processes to be investigated separately across covariates was taken up in parameter-specific distance metric (PSDM) GWR by allowing the non-ED metrics from Lu *et al*. (2014a) to vary across the covariates. The PSDM GWR framework was then extended to also incorporate ideas from Lu *et al*. (2016) by allowing parameter-specific Minkowski distances to be derived from the data again at the cost of significant computational overhead (Lu *et al*. 2018). Minkowski GWR and PSDM GWR are typically shown to achieve slight improvements in model fit compared to a traditional GWR using ED or non-ED metrics (Lu *et al*. 2017) or MGWR (Lu *et al*. 2018), respectively. However, there are several issues with these studies we believe readers should be aware of that make it difficult to confirm whether or not the gains of Minkowski GWR and the PSDM GWR framework are worth the increased computational time and potential loss of model interpretability. These issues are addressed in detail below.

3. Issues

3.1. *Veracity and importance of the results*

The assessment of the PSDM GWR framework is obscured by several points of contention. First, claims that non-ED and PSDM GWR offer worthwhile performance gains is exaggerated. This is due to how the results are reported and compared. For example, some of the figures reporting parameter estimates for different model specifications have axes with very small ranges that are not conducive to the metric being reported. For instance, Figure 1 in Lu *et al.* (2018, p. 52) presents the parameter estimate surfaces' root mean square error (RMSE) values across a series of models with varying Minkowski distances (i.e. $p = 0, 1, 2, \infty$) calibrated on simulated data using three known synthetic parameter surfaces (i.e. $\beta_0, \beta_1, \beta_2$) with varying levels of spatial heterogeneity. Since the RMSE is an un-normalized measure of error, its magnitude depends on the size of the values being compared. Examining the figure in Lu *et al.* (2018), it can be seen that the ranges of the RMSE values for each subplot ($\beta_0 \cong 0.023, \beta_1 \cong 0.01, \beta_2 = 0.015$) are all relatively small compared the magnitude of the known parameter values ($\beta_0 = \{3\}, \beta_1 = \{1, 5\}, \beta_2 = \{1, 5\}$). Therefore, despite the patterns in RMSE displayed in the figures, the RMSE and changes within the RMSE are all quite small relative to the estimated effects.

Another example of obfuscation stems from inconsistent model comparisons, such as when model fit improvements are sometimes compared against OLS results and sometimes compared against competing GWR specifications. Table 1 contains the R^2 model fit metrics reported throughout the non-ED and PSDM GWR literature for the same model specification for the LHP modeling example (Lu *et al.* 2014a, 2016, 2017, 2018). In addition, the percentage change in model fit is calculated using the R^2 from the OLS or standard GWR models as a baseline. It can be seen that using the two different baselines for comparing model fit improvements produces values with much different magnitudes. Thus, in order to better understand the relative performance increases amongst GWR specifications it is necessary to consistently compare the results to a common baseline result. Moreover, using a common baseline shows that using non-ED or TT for GWR or allowing for different distance metrics in multiscale models provides relatively small gains in model fit compared to standard GWR or MGWR for this example. This means that the differences between the GWR specifications in this study are minor and it is not clear that the more complex specifications are worthwhile. It would be insightful to carry out the same comparative exercise using the residual sum of squares (RSS) and the AICc, but these values are either not consistently reported across the aforementioned literature or are not calculated correctly. This later issue is discussed next.

A second point of contention is that different generalized additive models (GAMs) (i.e. PSDM GWR and MGWR) are evaluated by assigning the minimum component-wise AICc across all iterations during the backfitting calibration process as a descriptor of the overall model fit (Lu *et al.* 2017, 2018). Such a criterion is not a well-defined measure of

Table 1. Model fit reported in the literature for different models calibrated on the London house price data for the same specification according to R^2..

Specification: $PURCHASE_i = \beta_{0i} + \beta_{1i}FLOORSZ_i + \beta_{2i}PROF_i + \beta_{3i}BATH2_i$								
	OLS	GWR(ED)	GWR(ND)	GWR(TT)	FB-ED-GWR	FB-TT-GWR	BP-PSDM-GWR	PSDM
R^2	0.708	0.864	0.875	0.885	0.91	0.905	0.911	0.901
$\% R^2_{(OLS)}$	-	18.055	19.086	20	22.198	21.768	22.283	21.421
$\% R^2_{(GWR)}$	-	-	1.257	2.373	5.055	4.53	5.159	4.107

ED = Euclidean distance, ND = Non-Euclidean distance, TT = travel time, FB = flexible bandwidth, BP = best performing, PSDM = parameter-specific distance-metric.

model performance since the model fit for each component of the GAM is based on a regression of the partial residuals from the previous step of the calibration algorithm. That is, the component-wise AICc values are not computed from a proper overall hat matrix that projects the values of the response variable onto the final predicted values. Our results using the LHP data (Table 2) and the PBA data (Table 3) show that the appropriate AICc value for the entire model, which requires a few full-model statistics developed in Yu *et al*. (2018), is not equivalent to any of the ad-hoc minimum component-wise AICc values. Therefore, some doubt should be cast on the results reported in Lu *et al*. (2017) and Lu *et al*. (2018) where the ad-hoc model fit criterion is employed to evaluate alternative model specifications in comparative experiments.

Third, the choice of distance metric is largely inconsequential compared to the selection of the correct optimized parameter-specific bandwidths, which is the objective of MGWR as described by Fotheringham *et al*. (2017). The results in Lu *et al*. (2017, 2018) and Fotheringham *et al*. (2017) reinforce this argument – the differences in parameter estimate surfaces caused by the use of different metrics are minor whereas using different bandwidths (and holding distance metric constant) can cause major differences in parameter estimate surfaces. Lu *et al*. (2017, p. 992) demonstrates this for the LHP data: the left column displays relatively strong linear relationships between GWR models that use exclusively either ED or TT while the middle and right columns demonstrate relatively weak or non-existent linear relationships between results from either GWR-ED or GWR-TT and PSDM GWR, respectively.

Fourthly, it would appear that the results presented from the PSDM GWR framework (Lu *et al*. 2017, 2018) are based on unstandardized variables and as such, the resulting parameter-specific bandwidths may not necessarily be comparable. In the context of MGWR, Fotheringham *et al*. (2017) describe how if the covariates and dependent variable are not on the same scale then it is possible for the process-specific bandwidths to be influenced by the range and variation of each covariate. Indeed, in other local spatial models, variance and bandwidth are known to be inseparable (Warnes and Ripley 1987, Zhang 2004). To demonstrate the importance of this issue, the estimated bandwidths using standardized and unstandardized variables for the LHP data (Table 4) and PBA data (Table 5) were generated using MGWR. It can be seen that for most covariates a different bandwidth is obtained and in some cases the difference is relatively large (i.e. Intercept in Table 4 or Review Score in Table 5). The difficulty of interpreting the bandwidths in comparison to each other is compounded if different distance metrics

Table 2. Bandwidths and AICc for MGWR and PSDM GWR calibrated on the London house price data for the same specification.

Specification: $PURCHASE_i = \beta_{0i} + \beta_{1i}FLOORSZ_i + \beta_{2i}PROF_i + \beta_{3i}BATH2_i$					
MGWR	Entire Model	Intercept	FLOORSZ	PROF	BATH2
BW	-	46	18	242	84
AICc	368.722	-	-	-	-
PSDM	Entire Model	Intercept	FLOORSZ	PROF	BATH2
BW	-	47	18	154	84
AICc	-	242.389	291.898	214.702	225.209

All variables have been standardized to have a mean of 0 and variance of 1 and the minimum possible bandwidth is set to 2 nearest-neighbors. MGWR results pertain to the *mgwr* implementation while the PSDM results pertain to the *GWmodel* implementation.

Table 3. Bandwidths and AICc for MGWR and PSDM GWR calibrated on the Prenzlauer Berg AirBnB rental price data for the same specification.

Specification: $log(PRICE)_i = \beta_{0i} + \beta_{1i}REVIEWSCORE_i + \beta_{2i}ACCOMMODATES_i + \beta_{3i}BATHROOMS_i$					
MGWR	Entire Model	Intercept	Review Score	Accommodates	Bathrooms
BW	-	190	1279	79	2200
AICc	5245.832	-	-	-	-
PSDM	Entire Model	Intercept	Review Score	Accommodates	Bathrooms
BW	-	189	1279	79	2200
AICc	-	5075.262	5017.855	5162.689	5011.011

All variables have been standardized to have a mean of 0 and variance of 1 and the minimum possible bandwidth is set to 2 nearest-neighbors. MGWR results pertain to the *mgwr* implementation while the PSDM results pertain to the *GWmodel* implementation.

(i.e. meters vs. seconds) are used for each covariate and especially if the kernel used is a Gaussian one as in Lu *et al*. (2017). The bandwidth from a Gaussian kernel does not have a straightforward interpretation in comparison to the bisquare kernel where the bandwidth indicates the distance or number of neighbors at which data are weighted to zero (i.e. further observations have no influence). Therefore, comparing a bandwidth from ED and TT is particularly problematic in terms of deducing information on the relative spatial scales at which processes occur.

Finally, it is unclear whether or not the Minkowski approach used in both GWR (Lu *et al*. 2016) and PSDM GWR (Lu *et al*. 2018) chooses non-ED metrics because the ED metric is a misspecification or because the model is overfitting. Though Lu *et al*. (2016) perform simulation experiments, their experimental design does not generate data using known values of the Minkowski parameters, p and theta, and therefore does not properly verify the approach. Furthermore, their null model (scenario iii) is actually a scenario where both the data and the known processes surface are randomly distributed in space – a scenario where no variant of GWR should be seen as appropriate. More work is needed to show that the Minkowski approach recovers a known true distance metric alongside a known true bandwidth on that metric and is not instead convoluting different aspects of the

Table 4. Bandwidths for MGWR calibrated using the *mgwr* software on the London house price data for the same specification using standardized and unstandardized covariates.

Specification: $PURCHASE_i = \beta_{0i} + \beta_{1i}FLOORSZ_i + \beta_{2i}PROF_i + \beta_{3i}BATH2_i$				
MGWR(*mgwr*) BWs	Intercept	FLOORSZ	PROF	BATH2
Standardized	315	19	315	151
Unstandardized	46	18	242	84

All variables have been standardized to have a mean of 0 and variance of 1 and the minimum possible bandwidth is set to 2 nearest-neighbors.

Table 5. Bandwidths for MGWR calibrated using the *mgwr* software on the Prenzlauer Berg AirBnB rental price data for the same specification using standardized and unstandardized covariates.

Specification: $log(PRICE)_i = \beta_{0i} + \beta_{1i}REVIEWSCORE_i + \beta_{2i}ACCOMMODATES_i + \beta_{3i}BATHROOMS_i$				
MGWR(*mgwr*) BWs	Intercept	Review Score	Accommodates	Bathrooms
Standardized	190	1279	79	2200
Unstandardized	268	2110	129	2202

All variables have been standardized to have a mean of 0 and variance of 1 and the minimum possible bandwidth is set to 2 nearest-neighbors.

model. Given these concerns, it is not clear why the Minkowski approach is superior to using ED when there is no obvious theoretical non-ED metric.

3.2. *Prediction versus inference*

Another contention throughout the literature developing the PSDM GWR framework is that it is unclear whether the method is intended primarily for spatial prediction or also for exploratory/inferential work. While Lu *et al*. (2014a, 2016, 2017, 2018) discuss differences in parameter estimates across various model specifications, there is no mention of inference nor discussion of whether or not any framework is able to more reliably estimate parameter surfaces. When there are differences in parameters across specifications it is important to understand whether these differences are substantive or if they are statistically equivalent to noise. A GWR-specific hypothesis testing procedure has been developed (da Silva and Fotheringham 2016), explored (Fotheringham and Oshan 2016), and extended to MGWR (Yu *et al*. 2018) but to date similar frameworks have not been developed for non-ED GWR, Minkowski GWR or PSDM GWR. Rather, it seems that the PSDM GWR framework is aimed at spatial prediction[3] As such, it would be appropriate to adopt an out-of-sample prediction accuracy approach to evaluate performance across specifications while accounting for overfitting instead of the in-sample prediction accuracy that is currently reported in the literature (Hastie *et al*. 2009, James *et al*. 2013). Thus, the distinction between focusing on inference and prediction is important because each scenario suggests different criteria for evaluating competing specifications (Hofman *et al*. 2017).

3.3. *Tradeoff between interpretability and computational complexity*

Both MGWR and PSDM GWR are considered ‘big’ models because they estimate a large number of parameters that can be used to better understand spatial processes, but this advantage typically comes at the cost of increased computational overhead for model calibration. Evaluating whether or not the increased computational cost that comes with a new model specification is worthwhile is therefore an important area of study in the big model paradigm. Section 3.1 discusses several reasons why model fit may not be the best criterion to evaluate whether or not the additional computation time and complexity of PSDM GWR are worthwhile. A more important criterion for judging the merit of increased computation and complexity would be the extent that a specification lends itself to better substantive interpretation – a big model must be more meaningful to justify its expense. For example, MGWR will take longer to compute than GWR but the model interpretation is significantly enhanced if the bandwidths for each parameter are distinct. Hence, MGWR results are able to provide more accurate coefficient surface estimates that are distinct from standard GWR estimates and the multiple estimated bandwidths of MGWR can be interpreted as relative indicators of process scale for the relationship between each explanatory variable and the dependent variable (Fotheringham *et al*. 2017, Wolf *et al*. 2018, Oshan *et al*. 2018).

In contrast, using the data to choose (parameter-specific) distance metrics typically requires as much or more computation effort than manually selecting them, but the interpretability of the model could be compromised. In the best case scenario, the bandwidths for explanatory variables weighted with two different distance metrics can be transformed into common units in order to be compared, such as using average speed on a network to translate travel time to ED (Lu *et al*. 2014a, 2017). It is important to note that these conversions are approximate and in some cases it is not possible to translate between metrics, such as non-Euclidian distance and travel time where the appropriate translation is unclear. In the worst case scenario, when certain Minkowski metrics are selected, such as those associated with a fractional Minkowski norm (i.e. value of p), it is difficult, if not impossible, to intuitively interpret the estimated bandwidths in terms of spatial scale using standard concepts such as distance, nearest-neighbors, or travel time. Lack of interpretability is therefore a major drawback and may not be worth the additional computational costs or complexity even if the model fit is slightly improved from GWR or MGWR.

Figure 1 further demonstrates the interpretative value of ED over Minkowski metrics for GWR/MGWR. Here the bandwidth and *p* value estimates for ED GWR, Minkowski GWR, and ED MGWR using 100 random subsets of 85% of the LHP data are reported. It can be seen that GWR more consistently estimates the bandwidth than Minkowski GWR (left) and that the *p* value from Minkowski GWR (middle) appears to have little interpretative value because it spans the entire range of possible values (.25–8.0) across the 100 data samples. It can also be seen that MGWR further increases our ability to interpret scale (right): the bandwidth estimates for β_0 and β_1 are more local and more robust than those for β_2 and β_3 across the 100 calibrations. These results reaffirm the notion that data-driven distance metrics contribute less towards the interpretation of scale compared to MGWR using simple ED.

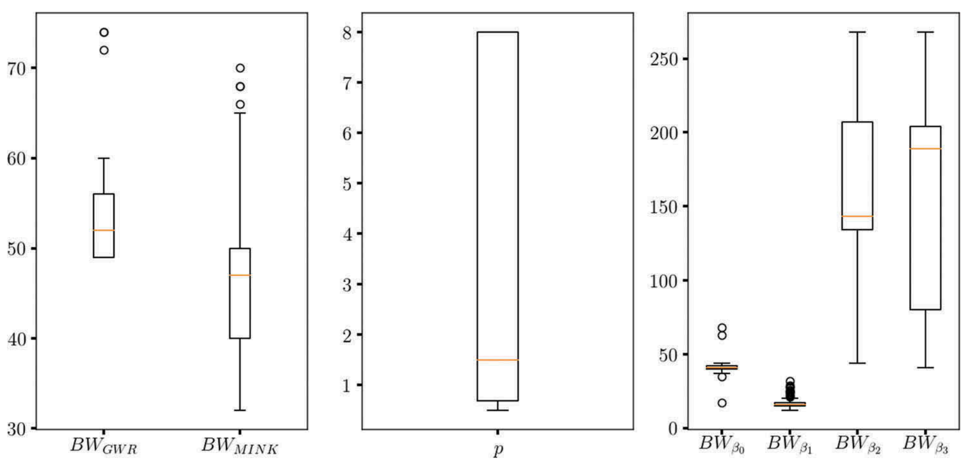

Figure 1. Boxplots of estimated bandwidths for GWR and Minkowski GWR (left), estimated Minkowski p values (middle), and estimated bandwidths for MGWR (right) for the same specification using 100 random subsets of 85% of the London house price data. The MGWR results are based upon standardized covariates. GWR and MGWR were calibrated using the mgwr software while Minkowski GWR was calibrated using GWmodel.

Though efforts have been made to reduce the computation time of PSDM GWR (Lu *et al*. 2016, 2017, 2018), the proposed heuristic shortcuts can result in sub-optimal solutions. The function for PSDM in *GWmodel* (Lu *et al*. 2014b) integrates some of these heuristics, which cannot be turned off, so it is hard to directly demonstrate this potential sub-optimality. One way around this is by calibrating a univariate[4] GWR using the Minkowski approach (i.e. exhaustive search) and comparing it to a univariate PSDM GWR, which should be equivalent, as is the case for GWR/MGWR (both produce a bandwidth estimate of 19 and an AICc of 7523.18 using *mgwr*). Results from this experiment using *GWmodel* are as follows: the heuristic PSDM GWR yields a bandwidth estimate of 12 and an AICc of 7561.918[5] while the exhaustive Minkowski GWR yields $p = \infty$, a bandwidth estimate of 19, and an AICc of 7521.233. This suggests the potential sub-optimality of the heuristic PSDM GWR, though it would also be worth further investigation in the multivariate context and using additional datasets. Otherwise, heuristic PSDM GWR should be avoided when the main goal is inference on spatial processes and an alternative and more parsimonious way to reduce computation time is to simply use MGWR and to manually specify parameter-specific distance metrics when they are theoretically appropriate.

4. Conclusion

Lu *et al*. (2017) state, 'FB GWR [or MGWR] is a special case of PSDM GWR, when only ED's are specified' and argue that PSDM GWR should always be the default specification when using GWR to study spatially varying relationships (Lu *et al*. 2018). We disagree with several aspects of this statement. First, MGWR is a special case of PSDM GWR only when the same distance metric, whether Euclidean or non-Euclidean, is employed. MGWR in Fotheringham *et al*. (2017) does not require Euclidean metrics, nor does it assume the metrics must be the same. If the analyst chooses to manually set parameter-specific distance metrics, it does not substantively change the model calibration routine nor increase computation time; this parameter-specific distance metric model would remain an instance of MGWR. One could equally just use parameter-specific kernel functions in MGWR rather than parameter-specific distance metrics. Both are just minor variations to the main theme of MGWR. Second, a more apt name for PSDM GWR would be PSDM MGWR since it always includes MGWR as a component of its calibration routine. Third, MGWR is a simpler specification than PSDM GWR and, consequently, we strongly recommend MGWR as the starting point for empirical applications.

As we have discussed throughout this comment, the PSDM extension to MGWR would appear to afford relatively small gains in model fit at the cost of potentially obfuscating aspects of model interpretation and increasing computational overhead. MGWR provides a straight-forward solution to providing covariate-specific indicators of process scale and enhancing model interpretation with limited additional computational cost. Recent work proposes an inferential framework for MGWR (Yu *et al*. 2018), making it an important tool for efficient and practical analysis of multiscale processes. Nevertheless, several steps could be taken to further strengthen our understanding of multiscale processes and improve the multiscale analytical tools described here. First, future work should give more attention towards ensuring consistent model comparisons that explicitly aim to assess inferential or predictive capabilities, since these tasks suggest different criteria (Hofman *et al*. 2017). Second, there is a great need to continue refining best practices for measuring and interpreting the concept of scale. For example, Wolf *et al*. (2018) raise the issue of bandwidth uncertainty in local multivariate statistical models and more work is still needed to fully understand how this uncertainty can be considered in GWR and MGWR. Finally, additional applications would help verify and affirm the methods discussed throughout this comment. Such advancements would vastly improve both intuition and technical capabilities for analyzing process heterogeneity and scale.

Notes

1. Version 2.0–6.
2. Version 2.0.1.
3. Lu *et al*. (2016) state, 'This drawback tends to make the Minkowski approach more suitable for prediction purposes with GWR' (p. 17). Lu *et al*. (2017) conclude, 'a PSDM GWR model can clearly improve GWR model performance in terms of GoF and prediction accuracy over a GWR model specified with EDs' (p. 15). Lu *et al*. (2018) note that, 'As expected the BP-PSDM-GWR model makes the most accurate predictions' (p. 8).
4. The LHP data was used for this with housing prices ('PURCHASE') being modeled as a function of floor size in square meters ('FLOORSZ') and without an intercept.
5. GWmodel does not report the selected p and theta for the heuristic PSDM GWR.

Acknowledgments

The authors would like to acknowledge the support of the U.S. National Science Foundation (NSF Grant 1758786) and the support of the National Social Foundation of China (Grants 71473008 and 13&ZD166).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation [1758786];National Social Foundation of China [13&ZD166,71473008].

ORCID

Levi John Wolf http://orcid.org/0000-0003-0274-599X

References

Carling, K., Han, M., and Håkinsson, J., 2012. Does Euclidean distance work well when the p-median model is applied in rural areas? *Annals of Operations Research*, 201, 83–97. doi:10.1007/s10479-012-1214-2

da Silva, A.R., and Fotheringham, A.S., 2016. The multiple testing issue in geographically weighted regression. *Geographical Analysis*, 48 (3), 233-247.

Fotheringham, A.S. and Oshan, T.M., 2016. Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems*, 18 (4), 303–329. doi:10.1007/s10109-016-0239-5

Fotheringham, A.S., Yang, W., and Kang, W., 2017. Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, 107 (6), 1247–1265. doi:10.1080/24694452.2017.1352480

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., and Harris, P., 2015. GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models . *Journal of Statistical Software*, 63 (1), 1–50.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The elements of statistical learning: data mining, prediction, and prediction*. 2nd ed. New York: Springer.

Hofman, J.M., Sharma, A., and Watts, D.J., 2017. Prediction and explanation in social systems. *Science*, 355 (6324), 486–488. doi:10.1126/science.aal3856

James, G., *et al*., 2013. *An introduction to statistical learning: with applications in R*. New York: Springer.

Jones, S., *et al*., 2010. Spatial implications associated with using Euclidean distance measurements and geographic centroid imputation in health care research. *Health Services Research*, 45 (1), 316–327. doi:10.1111/j.1475-6773.2009.01044.x

Lu, B., *et al*., 2014a. Geographically weighted regression with a non-euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, 28 (4), 660–681. doi:10.1080/13658816.2013.865739

Lu, B., *et al*., 2014b. The GW model R Package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science*, 17 (2), 85–101. doi:10.1080/10095020.2014.917453

Lu, B., *et al*., 2016. The Minkowski approach for choosing the distance metric in geographically weighted regression. *International Journal of Geographical Information Science*, 30 (2), 351–368. doi:10.1080/13658816.2015.1087001

Lu, B., *et al*., 2017. Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 31 (5), 982–998. doi:10.1080/13658816.2016.1263731

Lu, B., *et al*., 2018 April. Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. *Computers, Environment and Urban Systems*, 71, 41–57. doi:10.1016/j.compenvurbsys.2018.03.012

Oshan, T.M., *et al*., 2018 October. Mgwr: a python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *OSF Preprints*, 2. doi: 10.31219/osf.io/bphw9.

Phibbs, C., 1995. Correlation of travel time on roads versus straight line distance. *Medical Care Research and Review*, 52 (4), 532–542. doi:10.1177/107755879505200406

Warnes, J.J. and Ripley, B.D., 1987. Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika*, 74 (3), 640–642. doi:10.1093/biomet/74.3.640

Wolf, L., Oshan, T., and Fotheringham, A.S., 2017. Single and multiscale models of process spatial heterogeneity: single and multiscale models. *Geographical Analysis*, 50 (3), 223–246. doi:10.1111/gean.12147

Wolf, L. J., Oshan, T.M., and Fotheringham, A.S., 2018. Single and multiscale models of process spatial heterogeneity. *Geographical Analysis*, 50 (3), 223-246.

Yang, W. 2014. An Extension of Geographically Weighted Regression with Flexible Bandwidths. Thesis (PhD). University of St Andrews.

Yu, H., *et al*. (2019) Inference in multiscale geographically weighted regression, Geographical Analysis doi.org/10.1111/gean.12189

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99 (465), 250–261. doi:10.1198/016214504000000241