Fotheringham, A. S., Yu, H., Wolf, L. J., Oshan, T. M. and Li, Z. (2022) On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes. *International Journal of Geographical Information Science*, 36(8), pp. 1485-1502.
(doi: 10.1080/13658816.2022.2034829)

On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes

A. Stewart Fotheringham[1*], Hanchen Yu[2], Levi John Wolf [3], Taylor M. Oshan[4], Ziqi Li[5]

1. Spatial Analysis Research Center, School of Geographical Sciences and Urban Planning, Arizona State University, USA

2. Center for Geographic Analysis, Harvard University, USA

3. School of Geographical Sciences, University of Bristol, UK

4. Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, USA

5. School of Geographical and Earth Sciences, University of Glasgow, UK

Corresponding Author: A. Stewart Fotheringham (*stewart.fotheringham@asu.edu)*

Funding

This work is supported by the National Science Foundation with Award Number (#2117455)

Citation

Please cite this work as:

A. Stewart Fotheringham, Hanchen Yu, Levi John Wolf, Taylor M. Oshan & Ziqi Li (2022): On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2022.2034829

To link to this article: https://doi.org/10.1080/13658816.2022.2034829

On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes

Abstract

Models designed to capture spatially varying processes are now employed extensively in the social and environmental sciences. The main strength of such models is their ability to represent relationships that vary across locations through locally varying parameter estimates. However, local models of spatial processes also provide information on the nature of these spatially varying relationships through the estimation of a 'bandwidth' parameter. This paper examines bandwidth at a conceptual, operational, and empirical level within the framework of geographically weighted regression, one of the more frequently employed local spatial models. We outline how bandwidth relates to three characteristics of spatial processes: variation; dependence; and strength. *keywords: bandwidth; spatial scale; MGWR; spatially varying coefficients; spatial processes*

1. Introduction

Statistical models of locally varying processes, such as geographically weighted regression (GWR) and multiscale geographically weighted regression (MGWR), are commonly employed to provide information on spatially nonstationary relationships - those which reflect processes that vary over space. Often, in model comparison exercises, a set of local models is fit, and then the local parameter surfaces are compared (Waller et al. 2007; Wheeler & Calder 2007; Wheeler & Waller 2009; Finley 2011; Oshan & Fotheringham 2018; Wolf et al. 2018; Murakami et al. 2019). Generally, this is the main interest of local modeling: we fit *local models* because we are interested in the structure of *local trends*. Thus, most comparisons and examinations of local models generally compare how efficiently and accurately local parameters are estimated and how good predictions of the dependent variable are.

However, local models are "big models." They generate a vast amount of information, and although local parameter estimates and predictions form a major component of this, estimates of process scale that are obtained in the calibration of such models are also important (Oshan et al. 2019; Murakami et al. 2019). Most local models provide some estimated parameter which describes how localized the models are. In the case of GWR and MGWR (henceforth, we refer to both simply as GWR for convenience), this parameter is referred to as a 'bandwidth'. [1] The bandwidth parameter is usually interpreted as providing information on how "local" a process is. Generally speaking, if the bandwidth parameter is small[2], it indicates that a process is "local," and that parameters for specific locations can be optimally estimated by using information obtained from very close to those locations. In contrast,

[1] In Bayesian spatially varying coefficients models (SVCM) the parameter is referred to as a 'decay' (Finley et al. 2007) or a 'range' (Banerjee et al. 2004) parameter.

[2] While this requires us to assume that covariance kernels have been specified in a particular way, this assumption is not onerous. When specified as a "decay" kernel, large "decay" parameter values indicate locality. Defining the bandwidth as inversely-proportional to a decay parameter is sufficient. Either form has been used in Bayesian local models, but the bandwidth form is solely used in the GWR literature.

if the bandwidth parameter is large, then parameter estimation at each location requires information obtained from a much wider range of locations.

This interpretation of bandwidth is directly relevant to many research questions since it suggests that a single measure can be made of how spatially varying are process estimates. Application examples include the exploration of spatial process heterogeneity associated with health outcomes, mortality rates, air quality, voting behavior, and soil composition (Fotheringham et al., 2019; Comber et al., 2020; Cupido et al., 2020; Oshan et al., 2020; Fotheringham et al., 2021). Although the link between bandwidth and process variability has been established in local *aspatial* models (Fan and Gijbels, 1995; Ruppert et al. 1995), the situation in local *spatial* models is more nuanced for the following reasons:

> (i) In aspatial local modeling the focus is solely on establishing the nature of the spatially global conditional relationship between a dependent variable y and an independent variable x in attribute space whereas in spatial forms of local models nonstationarity is viewed as being caused by intrinsically different responses to the same stimulus at different locations due to geographical context. The difference becomes clear when x is constant: in local spatial modeling if x were constant over space, the relationship between y and x could still vary due to geographical context whereas in an aspatial application, local modeling would have no meaning if x were constant.
>
> (ii) In local spatial modeling, the relationship between bandwidth and process variation is controlled by factors other than process variation, such as process spatial dependency and process strength, and the interplay of these factors on the resulting bandwidth is unknown.

Consequently, exactly what determines the optimal bandwidth in GWR models of spatially varying processes and what it measures remain research questions which this paper attempts to answer. Specifically, we examine the sensitivity of reported bandwidth values in GWR models to three characteristics of processes:

> (i) The degree to which the processes reflected in local parameter estimates vary over space. When a process becomes increasingly global, the local parameter estimates reflecting this process will become increasingly uniform over space, and the resulting bandwidth will become larger. We measure this system characteristic by *local parameter variance.*
>
> (ii) The degree to which a process exhibits spatial dependency, as measured by the spatial dependency of the local parameter estimates. If a process varied randomly over space, the resulting bandwidth would be very large; increasing spatial dependency in the process would lead to smaller bandwidths. We measure this by *local parameter spatial dependency.*
>
> (iii) The strength of the process being measured, controlled by the error variance in the model. Processes which are very weak (with large model error) and where changes in a covariate magnitude have little impact on the magnitude of the dependent variable may exhibit different bandwidths from those which are strong, *ceteris paribus*. We measure this by *local parameter* strength.

To do this, we generate known parameter surfaces according to three different data generating processes (DGPs) with known formal properties: a Gaussian Random Field (GRF), a Simultaneous Autoregressive field (SAR), and a plasma fractal. Agreement in the conclusions drawn across the three DGPs reduces the possibility that the results are an artefact of any one DGP. We then use realizations from these processes in a local regression context in order to investigate what system properties have the most influence on the bandwidth estimates generated by GWR-type models.

First, we discuss the model specifications and DGPs. We then explain the interpretation of bandwidth in GWR as well as the parameters of each data generating process. Next, we examine the behavior of bandwidth estimates in GWR in terms of local parameter variance, local parameter spatial dependence and local parameter strength. We conclude by generalizing from the empirical patterns across DGPs to yield insights into the theoretical and empirical meaning of "bandwidth" in GWR models.

2. The concept of bandwidth as a measure of process scale

Scale is a complex and multifaceted concept in geography, but in spatial statistics, scale has generally become synonymous with *bandwidth*. An extension of well-known 1-dimensional nonparametric statistics, two perspectives on bandwidth are generally used. The first, focuses on bandwidth as an artifact *from estimation*, considering it as a distance that provides "a specification of neighborhood size" (Cleveland and Devlin, 1988 p. 597); points closer than the bandwidth are included in a local model, and points further than the bandwidth are not. This line of reasoning is well established among kernel methods in nonparametric statistics, where bandwidth only serves as a tuning parameter to improve the predictions of the model. There is no "true" value of bandwidth beyond that which optimizes a statistical score.

The second perspective generally comes *from data generating processes*. In this, bandwidth is an unknown parameter of the data generating process that establishes a "neighborhood size" for each site. For some DGP, then, the bandwidth parameter governs how far distant sites must be in order to be independent of one another. From this perspective, bandwidth has a "true" value in the DGP that an estimator ought to be able to uncover.

Practically speaking, this exposes an issue for the second perspective: how effective are our models at recovering the expected values of these bandwidths? Since the true structure of the DGP is unknown, it is generally not feasible to decide whether a given *process model* is correct for the problem at hand. As such, we would hope that bandwidth estimates generally have similar *semantics* between different DGPs, even if they arise from different estimators. Put simply, models should generally agree on what a "local" surface is in order for us to treat their bandwidth parameters as interchangeable.

However, interpreting bandwidth parameters is less straightforward than comparing the *response* surfaces, which typically have clear, direct meaning borrowed from a classic linear modelling case. Further, this is generally not a problem for aspatial local regression, since the bandwidth parameters used in that form of modeling generally are not interpreted as intrinsically meaningful (i.e., nonparametric kernel statistics). However, for spatial statistics, bandwidth matters as it is often interpreted as a proxy for the spatial scale over which a process is relatively uniform (i.e., an explicit characteristic of the DGP).

To formalize this idea, we define a classic local linear model as:

$$y_i = \beta_i x_i + \varepsilon_i \tag{1}$$

where y_i indicates a realization of the process at some position i, with observed data x_i, local effect β_i, and error term ε_i. In the sense of Cressie and Wikle (1993, p. 112), there are two data sources of variability (that for y and x), and two modeled sources of variability (that for ε_i and β_i). It is the variation of β_i that we refer to as *local parameter variance*, since it reflects the fundamental variation in the local estimates. The magnitude of ε_i reflects *process strength* and arises because we do not measure the outcome of $\beta_i x_i$ perfectly. The lower the value of ε_i, the greater is the strength of the relationship represented by the local parameter β_i. Finally, we define *local parameter spatial dependence* as arising from the structure of correlation between the local parameters at different sites, while the absolute variation is defined as process *parameter variance*.[3] Different processes for β_i will encode these sources of (co)variation differently, and different models that estimate bandwidth may combine these sources in different ways. Whether bandwidth *estimates* from GWR are equally sensitive to these three sources of variation remains unknown.

To address this, we discuss the nature of bandwidth in GWR. Previous research has shown how to calculate the uncertainty around each optimized bandwidth to establish confidence intervals and tests of the statistical significance of differences between bandwidths (Li et al. 2020). Therefore, we concentrate here on the interpretation of the bandwidth parameters reported in GWR (and hence in its multiscale version, MGWR) in relation to the impact that parameter variability, parameter dependence, and measurement variability may have on them. Before doing so, we first discuss the concept of bandwidth in the GWR framework.

2.1 The concept of 'bandwidth' in a geographically weighted regression framework

Suppose we have an area of interest that has *n* sites at which data are available for a set of variables that we assume to be associated in the following manner:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad (2)$$

That is, we believe that variations in x_1 and x_2 cause y to vary in predictable ways which can be measured by estimating the parameters β_1 and β_2. In order to calibrate this model, the usual procedure is to assemble data from the n sites, ignoring their locations, and to assume that the relationships represented by β_1 and β_2 are constant over the area of interest. This latter assumption is generally untested and not based on evidence. An alternative is to allow the possibility that the relationships in the model vary over space and to rewrite the models as:

$$y_i = \beta_{0i} + \beta_{1i} x_{1i} + \beta_{2i} x_{2i} + \varepsilon_i \qquad (3)$$

where now the parameters of the model can vary with location i. The problem with this approach is that traditionally it would need a relatively large number of observations on y, x_1 and x_2 to be observed at each site *i* whereas typically with spatial data only one set of y, x_1 and x_2 values are available at each location. GWR-related models solve this problem by calibrating the model separately at each location by borrowing data from neighboring locations and weighting them between 0 and 1 with greater weights being given to data from locations that are closer to that for which the local parameter estimates are being estimated. Figure 1 describes this process.

[3] Fundamentally, this is a partitioning of the covariance matrix of $\boldsymbol{\beta}_i$ into its diagonal and off-diagonal elements.

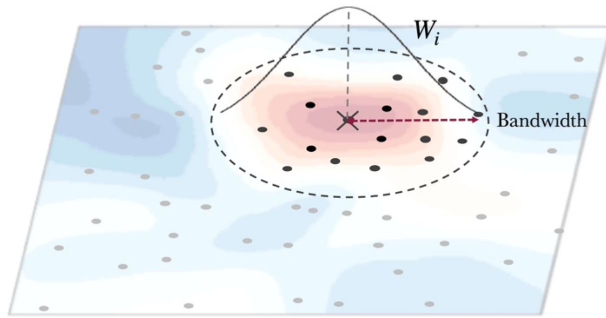

Figure 1: Data-weighting or borrowing in geographically weighted regression.

An integral component in this type of local model calibration is the function that relates the weight given to data from a certain location to the distance that location is from the regression point. In particular, the 'bandwidth' of a kernel function controls the decay of the weighting as distance increases. If the bandwidth is large, weights decline slowly as distance increases so that data from more sites are included in the local calibration; if the bandwidth is small, data from only a relatively small set of local sites are used in the calibration. The choice of optimal bandwidth is thus a trade-off between bias and variance. As more distant locations are added to the local regression, bias in the local parameter estimates will increase because data are drawn from further away where the relationships they represent are increasingly likely to be different to those at the regression point. However, adding more data to the local regression will decrease the standard errors of the local parameters as the local estimates will be based on greater numbers of observations (Yu et al. 2020a). The optimized bandwidth produced by a GWR calibration thus represents the number of locations (or distance from the regression point depending on what type of weighting function is used) at which adding a further set of data from the next closest location will increase the bias in the parameter estimates more than it will decrease the uncertainty about the estimates. Details on how to calculate the bias and uncertainty in GWR models can be found in Yu et al. (2020b).

Another way of viewing the bandwidth in GWR is via the concept of data-borrowing (Fotheringham and Sachdeva, 2021). Local regressions can be calibrated through GWR because data are 'borrowed' from nearby locations and weighted according to their proximity to the focus of the local regression. Each location can be thought of as providing data for the local regression at *i* which contains some information on the processes operating at *i* and some misinformation. All the data at locations up to the bandwidth provide more information than misinformation on the processes at *i* and are therefore used in the local regression. Data which lie beyond the bandwidth are composed of more misinformation than information on the processes at location *i* and are therefore excluded from the local calibration for location *i*.

In MGWR (Fotheringham et al. 2017), this bandwidth will vary according to the relationship between *y* and each covariate. Some relationships will have relatively small associated bandwidths, indicating that in such cases the amount of misinformation in data quickly exceeds the amount of information as distance from the local regression point increases. Other relationships will have relatively large bandwidths, indicating that levels of misinformation tend to increase very slowly as distance from the regression point increases so that larger numbers of data points are included in the local regressions to reduce parameter estimate uncertainty. Relationships that do not vary over space will have, in theory, infinitely large bandwidths as adding data introduces no bias (or 'misinformation') but will reduce the standard errors of the parameter estimates. Consequently, the resulting optimized bandwidths that result from a GWR calibration are of interest in that they would appear to suggest the spatial scale over which relationships vary. This hypothesis is examined below.

3. Experimental Design

To examine the extent that the optimized bandwidth parameters from GWR are related to various system features, we conduct a variety of simulations which are now described. Local parameter estimate surfaces are simulated using three different data-generating processes: a Gaussian Random Field, a Simultaneous Autoregressive Field, and a bounded discrete-space Fractal process. Each of these processes has a parameter governing how 'dependent' or 'local' their realizations become, as well as a parameter governing the 'intrinsic' variation of the surface. This allows us to examine the behavior of the bandwidth estimate from GWR for different kinds of processes, without reference to what the "correct" process specification might be (as has been investigated in Harris (2019)). The overall logic of the experiments is to generate surfaces of local parameter values with known properties of process variability, process spatial dependence and process strength and to uncover how these properties affect the resulting bandwidth parameters in GWR.

In all the simulations, we use univariate GWR to calibrate optimal bandwidths with different data that have known properties and which are generated by different processes. The model is kept as simple as possible because of the time-consuming nature of the simulations, yielding the situation where a variable y is a function of a single covariate x:

$$y_i = \beta_i x_i + \varepsilon_i \tag{4}$$

where values of the covariate x_i are randomly drawn from a normal distribution (0, 1) for 400 locations on a regular 20×20 grid, the local parameters β_i are obtained from a pre-specified surface, and ε_i are drawn from a normal distribution (0, 1). The values of y are then derived from the model equation. Though not pursued here, this experimental design could be extended to the multiscale version of GWR although some preliminary experiments with MGWR suggest the conclusions we draw here from simple GWR apply equally to the covariate-specific bandwidths derived in MGWR. Indeed, the model we employ here with only one covariate can be considered the simplest case of MGWR.

4. Bandwidth sensitivity in GWR to process variance and process spatial dependency

4.1 Bandwidths based on data derived from a Gaussian Random Field (GRF) process

For the first DGP, we use a mean-stationary Gaussian random field (GRF) to generate a set of local parameter surfaces where Ω denotes the covariance matrix arising from a squared exponential spatial covariance function as discussed above:

$$\Omega(k, h) = k^2 * exp(-0.5 * \frac{d_{ij}^2}{h^2}) \tag{5}$$

In this notation, h is the parameter controlling the surface's autocorrelation, k is the intrinsic/residual variation (sometimes called the "microscale variation" after Cressie (1993)), and d_{ij} is the distance between locations i and j. Thus, the spatial dependence between the local slopes increases when h increases. In addition, the strength of realizations of the local slopes will increase when k gets larger.

Thus, by systematically varying *k* and *h*, we can examine the resulting impacts on bandwidth estimates from local models.

To examine the sensitivity of the resulting optimized bandwidths to both dependency in the local parameters and their strength, we vary *h* from 0.2 to 10 in steps of 0.2 and *k* from 0.05 to 1 in steps of 0.05. By doing this, we obtain 1000 pre-specified local parameter surfaces from the GRF DGP with different degrees of dependency and strength. For each of these surfaces, we calculate the optimal bandwidth and Moran's I[4]. To give a feel for what these surfaces look like in terms of their measured spatial dependence and strength, six representative surfaces of parameters are shown in Figure 2. The equivalent surfaces for the subsequent two DGPs are very similar and are not included in the paper but are available from the authors.

[4] Here, we use binary first-order rook contiguity.

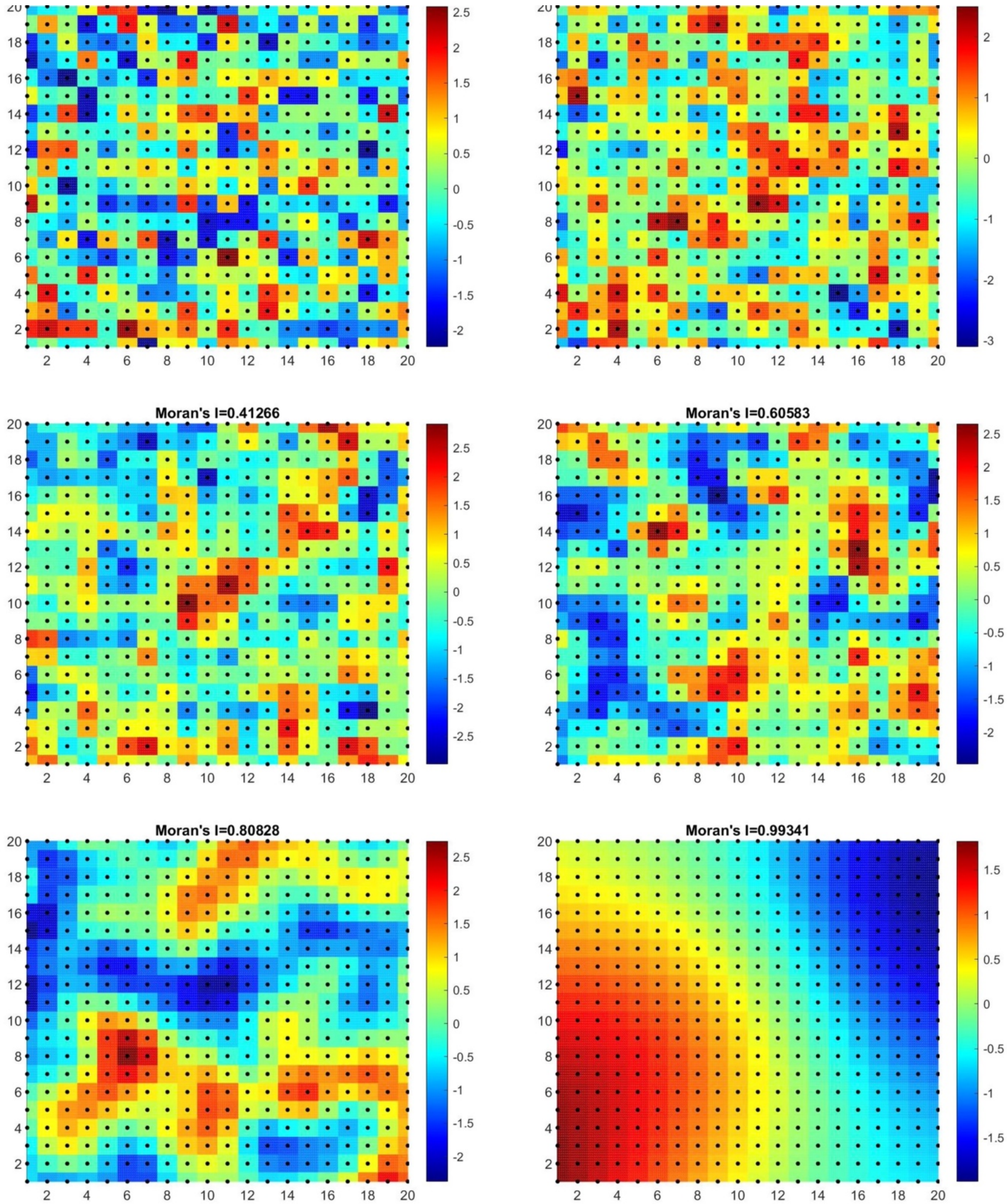

Figure 2: Six examples of parameter surfaces with an increasing degree of process spatial dependence based on the Moran's I statistic.

In order to remove the influence of randomness from the results, we produce 100 simulations for each of the 1000 combinations of dependency and strength and average the resulting optimized bandwidths for each combination. These averages are reported in Figure 3 where each cell denotes the optimal number of nearest neighbors used to calibrate a GWR model, which ranges from 5 (indicating a very localized process) to 400 (indicating a global process).

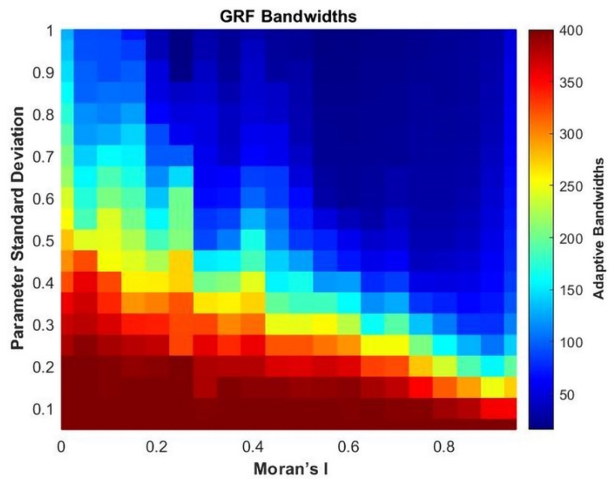

Figure 3: Optimal bandwidths in terms of number of nearest neighbors obtained at different levels of local parameter variance (vertical) and local parameter spatial dependence (horizontal) averaged over 100 samples derived from a GRF.

The figure shows that when parameter spatial dependency (as measured by Moran's I) is held constant, bandwidth decreases as the variation in the local parameters increases and that when the variation in the local parameters is held constant, bandwidth decreases as spatial dependency increases, although the sensitivity is not as strong in this latter case. The largest bandwidths are generally found in situations where the parameter surface exhibits low variance. When there is little variation in a process over space, the optimal bandwidth will be large, *ceteris paribus,* because using data from locations further away in the local regressions introduces little bias. In contrast, the smallest bandwidths are generally found when the parameter surface exhibits high variance. With fixed spatial dependency, as the process exhibits increasing spatial variation, the optimal bandwidth will decrease as using data from more distant locations will introduce more bias. This result is further highlighted in Figure 4 where slices across the horizontal axis at several values of parameter standard deviation are taken from Figure 3 and plotted as lines. Overall, Figure 4 demonstrates that larger bandwidths are obtained when parameter variation is low and smaller bandwidths are obtained when parameter variation is high.

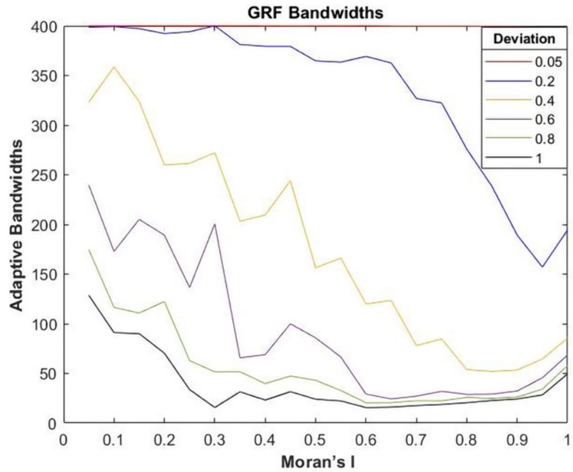

Figure 4: Optimal bandwidths for six fixed values of parameter surface variation and varying spatial dependency.

These results match the intuition that small bandwidths are associated with large variations in parameters (processes), reflecting conditioned associations between a covariate and the dependent variable that vary rapidly over space. In such situations, borrowing data from more distant locations would yield high levels of bias - the amount of 'misinformation' in data quickly exceeds the amount of 'information' as distance from the regression point increases. When there is little or no variation in parameters (processes), there is nothing to be gained from a small bandwidth - data can be borrowed from locations at much greater distances which reduces parameter uncertainty. So, small bandwidths are associated with processes that vary rapidly over space; large bandwidths are associated with processes that are relatively constant over space. The relationship between bandwidth and parameter (process) variation is somewhat tempered by the degree of spatial dependency exhibited by the parameters (processes). Holding parameter variation constant, optimal bandwidth tends to decline slightly as spatial dependency increases. This also is intuitive because if processes were randomly distributed over space, data would be borrowed from more distant locations to reduce parameter uncertainty. As parameters (processes) exhibit greater spatial dependency, optimal bandwidths decrease to reduce the amount of bias introduced by including more distant (and more different in terms of processes) locations in the local calibrations. Therefore, optimal bandwidths are strongly related to the amount of parameter variation but only when there is some degree of spatial dependency in the processes reflected by the parameters. Given that almost all spatial data exhibit some degree of positive spatial dependency, it seems reasonable to assume that spatial processes, if they vary, will also exhibit some degree of spatial dependency. In addition, there are few empirical examples of spatial processes in the social sciences that vary randomly over space, lending further credence to the assumption that spatial process variation and dependence tend to coexist[5] .

[5] While some spatial processes related to physical phenomena, such as soils, appear to be 'random', this is often an issue of data and scale, rather than the true underlying process itself.(Webster, 2000)

4.2 Bandwidths based on data derived from a Spatially Autoregressive (SAR) process

While the GRF DGP used to generate the parameter surfaces in Section 3.1 is common in spatial statistics, other DGPs are also used to model spatially-dependent phenomena. To test whether the results from the previous experiment are generalizable, we repeat the above experiment with two additional data-generating processes. First, we assume the pre-specified surfaces S~SAR(ϱ, k) follow a simultaneous autoregressive process (SAR) which is completely characterized by its spatial lag component ϱ and its standard deviation k (Anselin 1988). A SAR process is generated as follows:

$$S = (I - \rho W)^{-1} \varepsilon \tag{6}$$

where ε is drawn from a normal distribution $(0, k^2 I)$ and W is a rook contiguity weights matrix. The dependency of the surface increases as ϱ becomes larger while the variance of the surface increases as k becomes larger.

To investigate the sensitivity of the resulting optimized bandwidth to variations in both ϱ and k, we vary ϱ from 0 to 0.95 in steps of 0.05 and k from 0.05 to 1 in steps of 0.05. This generates 400 surfaces of parameter values with different degrees of dependency and variance and for each of these we run 100 simulations and obtain the average optimized bandwidth over these 100 simulations, which are reported in Figure 5.

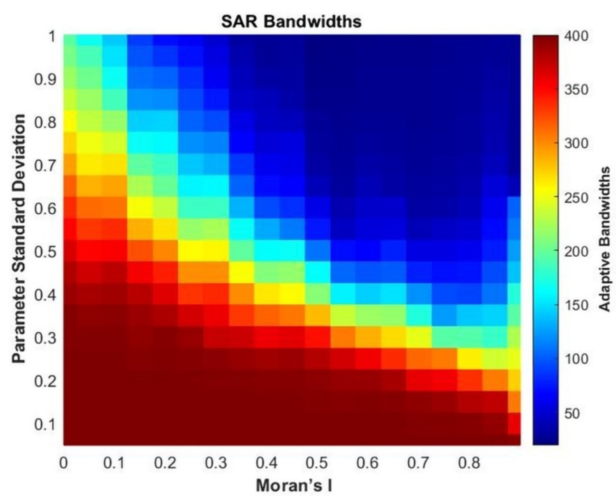

Figure 5: Optimal bandwidths in terms of number of nearest neighbors obtained at different levels of local parameter variance (vertical) and local parameter spatial dependence (horizontal) averaged over 100 samples derived from a SAR process.

Reassuringly, the results are similar to those obtained under surfaces generated by the GRF DGP, although the influence of the spatial dependence of the parameters is greater in this case. Smaller bandwidths are associated with local parameter surfaces with higher spatial dependency and greater variation while larger bandwidths are associated with surfaces having lower spatial dependency and lower variation. The sensitivity of the optimized bandwidth to variations in process spatial dependency is greatest when process variance is moderate. Similarly, the sensitivity of the optimized bandwidth to process variation is greatest when process spatial dependency is not extreme. The tradeoff between process strength and process spatial dependence becomes clear: a "local" bandwidth arises in situations with large parameter variance or strong spatial dependence.

4.3 Bandwidths based on data derived from a Fractal process

Finally, we further reinforce our results by assuming that the pre-specified parameter surfaces S ~ Fractal(f, k) follow a plasma fractal process (Fournier et al., 1982; Willemse & Hawick, 2013) which is completely characterized by its roughness f and its standard deviation k. The dependency of the surface decreases as f becomes larger while the magnitude of the surface increases as k becomes larger. In order to investigate the sensitivity of the optimized bandwidth to variations in both spatial dependency and process strength as above, surfaces are created by varying ϱ from 0 to 2 in increments of 0.05 and k from 0.05 to 1 in steps of 0.05. This generates 820 pre-specified surfaces with different combinations of spatial dependency and process strength. The average optimized bandwidths based on 100 simulations for each combination are shown in Figure 6.

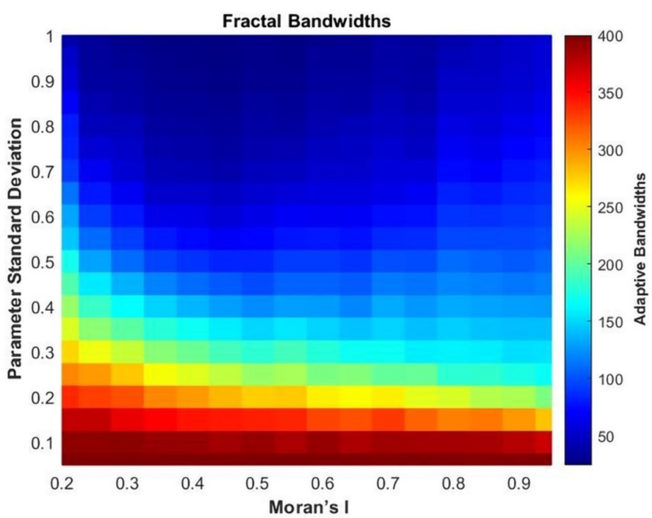

Figure 6: Optimal bandwidths in terms of number of nearest neighbors obtained at different levels of local parameter variance (vertical) and local parameter spatial dependence (horizontal) averaged over 100 samples derived from a fractal generating process.

Again, similar relationships are observed as those from the previous two experiments: bandwidth decreases with increasing parameter variation and increasing spatial dependency, though the sensitivity of bandwidth size is dominated more by parameter variation than spatial dependence. This supports the results obtained for the GRF and SAR DGPs but provides an example where the influence of spatial dependence on the bandwidth is even weaker. This could be due to the scale-free nature of fractal processes that maintains structural relations across scales or sizes of values in a set. Also note that the plot has a minimal dependence of 0.2 on the X-axis, which could skew interpretation.

From the above three experiments, it becomes clear that the optimized bandwidth in GWR is sensitive to the variation in the relationships represented by the local parameters and the degree of spatial dependency in relationships represented by the local parameters. Where there are highly spatially localized relationships, we would expect to generate small optimized bandwidths and where there are relatively constant relationships, it is expected that GWR will generate large bandwidths. Consequently, any optimized bandwidth generated in a GWR (or MGWR) calibration can be thought of as an indicator of the spatial scale over which a relationship varies.

5. Bandwidth sensitivity in GWR to variations in process strength (model error)

Having established a link between the optimized bandwidth and both spatial variation and spatial dependency in a given process, we examine here how variation in the strength of a relationship (as controlled by the error variance in equation (1)) affects bandwidth estimates. Overall, as the error variance in equation (1) decreases, the random component of y decreases. making the local slopes increasingly important in determining y at a fixed value of x.

In the GRF, we set k=1 and vary h in equation (5) from 0.2 to 10 in increments of 0.2, generating 50 pre-specified surfaces with different degrees of spatial dependency but with the same variance. For each of these 50 values of h we change the standard deviation of the error term in the model (sigma) from 0.1 to 4 in steps of 0.2. The optimal bandwidth and Moran's I for each surface is then calculated and the results are averaged across 100 realizations for each combination of h and sigma, which are displayed in Figure 7.

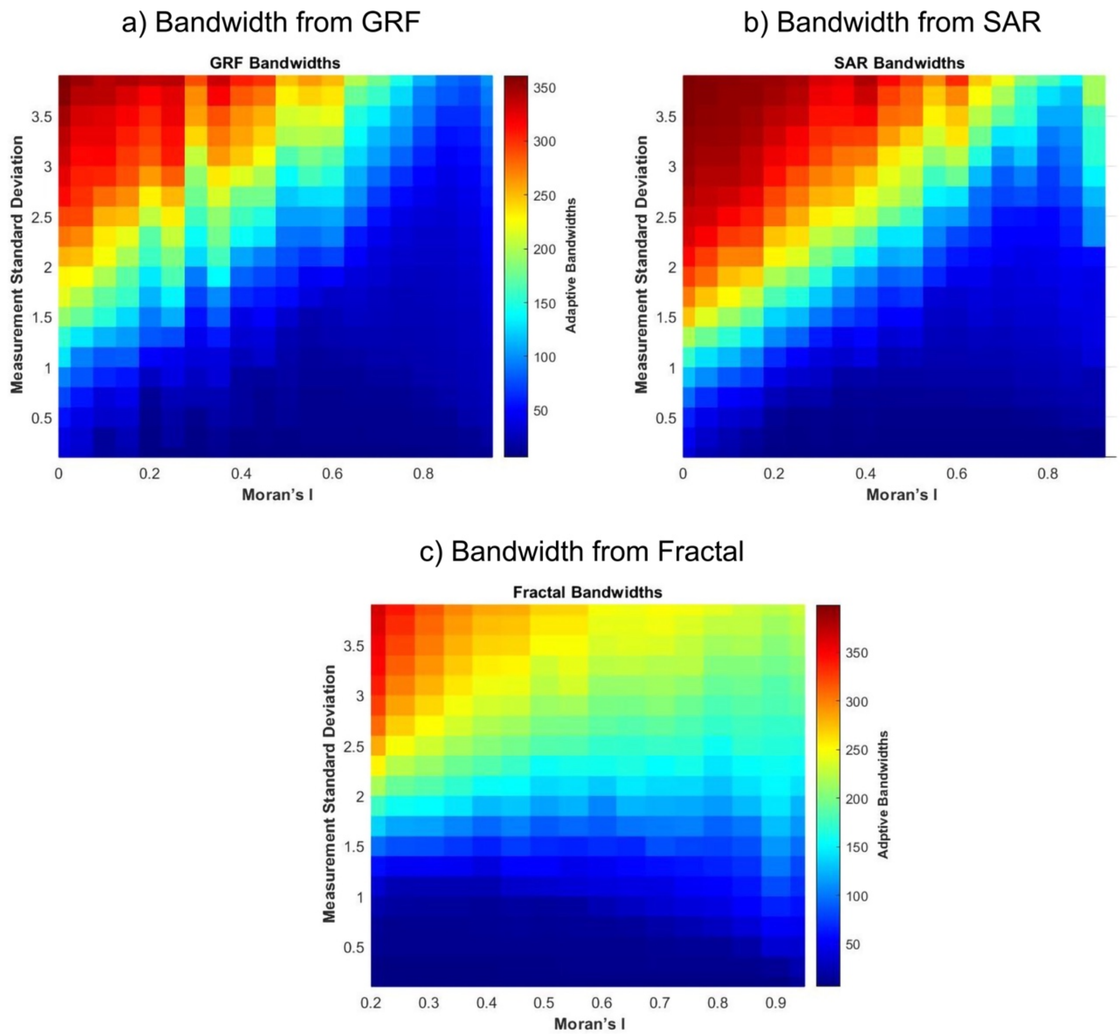

Figure 7: Optimal bandwidths in terms of number of nearest neighbors obtained at different levels of measurement error (vertical) and local parameter spatial dependence (horizontal) averaged over 100 samples derived from different processes (a) GRF; (b) SAR; and (c) Fractal.

The results indicate that when spatial dependence is held constant, bandwidth increases as the variance of the error term (i.e., noise) increases and process strength decreases. Recall that the optimized bandwidth is a tradeoff between bias and variance (i.e., parameter estimate standard error), such that as the noise increases (greater randomness and weaker process strength), the optimal bandwidth increases to include more data in the local regression, reducing parameter estimate uncertainty. This effect is most clearly seen when spatial dependency is low and diminishes as spatial dependency in the parameter surface increases. Furthermore, when spatial dependency is very high, the optimized bandwidth is typically small because local regression with only a relatively small number of data points produces parameter estimates with both low bias and variance.

The results in Figure 7 also indicate that if the level of noise in the model is held constant, the optimized bandwidth decreases as spatial dependency increases. This relationship is most clear for higher levels of noise and lower levels of spatial dependence, causing the optimized bandwidth to become very large and approach that of a global model. In contrast, when the noise in the model is consistently low, the optimized bandwidth will tend to become small because adding more data may increase bias without substantially reducing variance. This increase in bias becomes more substantial as spatial dependence increases.

6. Sensitivity to the Distribution of Locations

In order to check if the results described in this paper might be sensitive to the distribution of the locations at which data are recorded, we performed the following experiment. We selected one surface from Figure 7a where the optimal bandwidth is 50, I=0.98 and measurement error is 1.0. This surface is shown in Figure 8a. We then randomized the locations for which the data were recorded 100 times and for each randomized set of locations we recorded the optimal bandwidth. The frequency distribution of these 100 values is shown in Figure 8b.

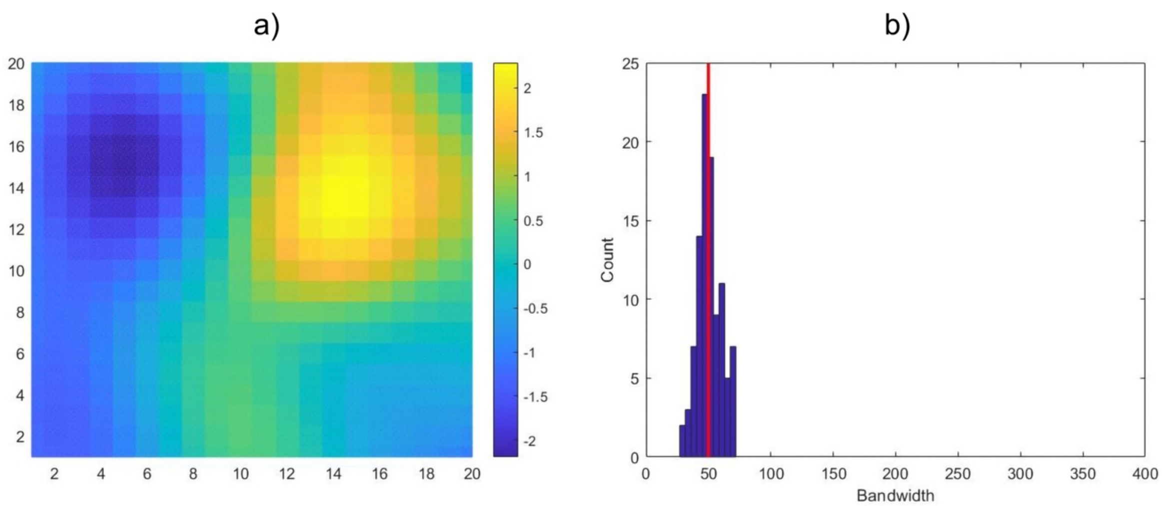

Figure 8: (a) Surface from Figure 7(a) with optimal bandwidth = 50, I = 0.98 and measurement error = 1; (b) Frequency diagrams showing 100 optimal bandwidths from randomizing the locations used to derive the optimal bandwidth in Figure 8(a).

The results from this single experiment demonstrate that the optimal bandwidths reported in the experiments described in section 5 would appear to be robust to variations in the distributions of the locations at which data are recorded. This gives us confidence that the results in the remainder of the paper are equally unaffected by the distribution of locations.

7. Discussion and Conclusions

This paper examines the nature of the bandwidth parameter in the Geographically Weighted Regression framework to clarify its interpretation as an indicator of the geographical scale over which a process varies. In the GWR framework, the bandwidth can be measured as a physical distance, although it is more usually measured in terms of the number of nearest neighbours around a local regression point from which data are weighted and borrowed to enable the calibration of a model at each regression point. In MGWR, the number of nearest neighbors that data are borrowed from can vary across the covariates in the model, providing information on how each separate process (conditional relationship between *y* and a particular *x*) varies over space. Each covariate-specific bandwidth in MGWR is a trade-off between bias and variance in the local parameter estimates for that covariate (Yu et al. 2020b). The bandwidth is the n^{th} nearest distance-weighted data point included in each local regression; data points beyond this threshold are weighted to zero. Beyond the bandwidth, the inclusion of further data points increases the bias in the local parameter estimates more than it reduces their uncertainty. Consequently, the bandwidth parameter is an important output in GWR and MGWR because it describes how local or global each process is being modeled. Small bandwidths indicate processes that vary rapidly over space; large bandwidths indicate processes that vary slowly over space. Both models have the global model as a special case when the bandwidth increases towards infinity. However, despite its importance as a potential indicator of process scale, very little is known about the determinants of the optimal bandwidth. Three factors related to process characteristics were theorized to potentially influence the optimized bandwidth parameter estimated from local models: (i) the degree to which processes vary over space, denoted by *local parameter variance*; (ii) the degree to which there is spatial dependency in processes, *measured by a Moran's I coefficient*; and (iii) the strength of the process, *measured by the error variance in the model* and which is inversely related to process strength.

The results from the experiments provide evidence that small bandwidths are indicative of processes that: (i) have a high degree of spatial variability; (ii) a high degree of spatial dependency; and (iii) are strong, in the sense that the covariate has a meaningful impact on the dependent variable. The dominant controlling factor on the bandwidth appears to be process variability but this is tempered by low process spatial dependence and low process strength (i.e., high noise). This means that a larger bandwidth could be reported, even though there is a moderate level of process variability if, for instance, there is weak process spatial dependence - indeed, if processes varied randomly over space, then there would be little to be gained from borrowing data from only nearby locations. On the contrary, large bandwidths are indicative of processes that are a) relatively stationary over space; b) nonstationary but have very weak spatial dependence; c) have a relatively weak relationship with the dependent variable; or d) some combination of thereof.

Although not reported in the main body of the paper, supplemental material below describes some preliminary research that suggests that the above results on the determinants of the bandwidth parameter within a GWR framework are also found within the Bayesian framework of a Spatial Gaussian Process model. Our findings suggest that the meaning of the bandwidth parameter remains the same in both specifications: Spatial dependence in a SAR process is captured in the bandwidth in the same fashion as that in a Fractal or GRF process for both kinds of local models.

In the main, this implies that differences in *system articulation* (Bivand 2017 citing Wilson 2000) — the step of analysis *before* model estimation where the analyst decides what effects to represent, how, and with what parameters—are not as critical for deciding between local model specifications. Even though they "use" bandwidth differently or may specify a different theoretical structure from the "true" data generating process, the meaning of their estimates is largely the same. This agreement will probably remain a feature of this literature. Instead, decisions should consider the practical differences between the two specifications.

Data and Code Availability Statement

The data and code that support the findings of this study are openly available in figshare at https://www.doi.org/10.6084/m9.figshare.14340368.

References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). Hierarchical Modeling and Analysis for Spatial Data. Boca Raton, FL: CRC Press.

Bivand, R. (2017). Revisiting the Boston data set: changing the units of observation affects estimated willingness to pay for clean air. *Region: the journal of ERSA*, *4*(1), 109-127.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, *83*(403), 596-610.

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., & Harris, P. (2020). The GWR route map: A guide to the informed application of Geographically Weighted Regression. *ArXiv:2004.06070 [Stat]*. http://arxiv.org/abs/2004.06070

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons.

Cressie, N., & Wikle, C. K. (1993). *Statistics for spatio-temporal data*. John Wiley & Sons.

Cupido, K., Fotheringham, A. S., & Jevtic, P. (2020.). Local modelling of U.S. mortality rates: A multiscale geographically weighted regression approach. *Population, Space and Place*.

Datta, A., Banerjee, S., Finley, A. O., & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, *111*(514), 800-812.

Fan, J., & Gijbels, I. (1995). Data‐driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(2), 371-394.

Finley, A. O. (2011). Comparing spatially‐varying coefficients models for analysis of ecological data with non‐stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, *2*(2), 143-154.

Finley, A. O., Banerjee, S., & Carlin, B. P. (2007). spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of statistical software*, *19*(4), 1.

Fotheringham, A S., C. Brundson, and M. Charlton. 2004. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. New York: Wiley.

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247-1265.

Fotheringham, A. S., Yue, H., & Li, Z. (2019). Examining the influences of air quality in China's cities using multi-scale geographically weighted regression. *Transactions in GIS*, 23(6), 1444–1464.

Fotheringham, A. S. and M. Sachdeva (2021) Modeling Spatial Processes in Quantitative Human Geography. *Annals of GIS*, DOI: 10.1080/19475683.2021.1903996

Fotheringham, A. S., Li, Z., & Wolf, L. (2021). Scale, Context and Heterogeneity: A Spatial Analytical Perspective on the 2016 US Presidential Election. *Annals of the American Association of Geographers.*

Fournier, A., Fussell, D., & Carpenter, L. (1982). Computer rendering of stochastic models. *Communications of the ACM*, *25*(6), 371–384.

Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*(462), 387-396.

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

Kaufman, C. G., Schervish, M. J., & D. W. Nychka. (2008). Covariance Tapering for Likelihood-based estimation in large spatial datasets. *Journal of the American Statistical Association,* 103(484): 1545-1555.

Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. *International Journal of Geographical Information Science* 34 (7): 1378–1397.

Li, Z., Fotheringham, A. S., Oshan, T. M., & Wolf, L. J. (2020). Measuring Bandwidth Uncertainty in Multiscale Geographically Weighted Regression Using Akaike Weights. *Annals of the American Association of Geographers.* 110(5): 1500–1520.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. Geographical Analysis, 32(1), 19-35.

Murakami, D., & Griffith, D. A. (2015). Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems*, *17*(4), 311-331.

Murakami, D., Lu, B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., & Griffith, D. A. (2019). The importance of scale in spatially varying coefficient modeling. *Annals of the American Association of Geographers*, *109*(1), 50-70.

Oshan, T. M., & Fotheringham, A. S. (2018). A Comparison of Spatially Varying Regression Coefficient Estimates Using Geographically Weighted and Spatial-Filter-Based Techniques: A Comparison of Spatially Varying Regression. *Geographical Analysis,* 50(1), 53–75.

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, *8*(6), 269.

Oshan, T. M., Smith, J. P., & Fotheringham, A. S. (2020). T*argeting the spatial context of obesity determinants via multiscale geographically weighted regression*. International Journal of Health Geographics, 19(1), 11.

Ruppert, D., Sheather, S. J., & Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, *90*(432), 1257-1270.

Wheeler, D. C., & Calder, C. A. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, *9*(2), 145-166.

Wheeler, D. C., Páez, A., Spinney, J., & Waller, L. A. (2014). AB ayesian approach to hedonic price analysis. *Papers in Regional Science*, *93*(3), 663-683.

Wheeler, D. C., & Waller, L. A. (2009). Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems*, *11*(1), 1-22.

Willemse, J. M., & Hawick, K. A. (2013). Generation and Rendering of Fractal Terrains on Approximated Spherical Surfaces. *Proceedings of The 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing*, WorldComp: Las Vegas, NV .

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

Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2020a). Inference in multiscale geographically weighted regression. *Geographical Analysis*, 52(1), 87-106.

Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., & Wolf, L. J. (2020b). On the measurement of bias in geographically weighted regression models. *Spatial Statistics*, 38, 100453.

Supplemental Material 1: Model Error behaves similarly between Spatial Gaussian Process and Multiple Bandwidth GWR Models.

In a similar fashion to the results shown in Section 6 of the main paper, Figures S1 and S2 show the effect of model error (top row = 0.4; bottom row = 1.0) for all DGPs on the bandwidth estimate from SVCM and GWR, respectively, holding parameter variability constant at 1. Again, both sets of results are roughly the same: because process variation is very high (set at 1.0 in all experiments), bandwidths tend to be small and get smaller as process spatial dependency increases. However, this relationship becomes weaker as measurement error increases (proces strength weakens). Increasing model error increases the required strength of process spatial dependence to detect a "local" process. The upturn in values observed for the GWR bandwidths at extreme levels of process dependency is probably caused by the parameter surfaces at such extreme levels of dependency being essentially planes with constant gradient and the optimized bandwidths will be very sensitive to minor variations in the surface, somewhat akin to the effect of extreme collinearity on parameter estimates.

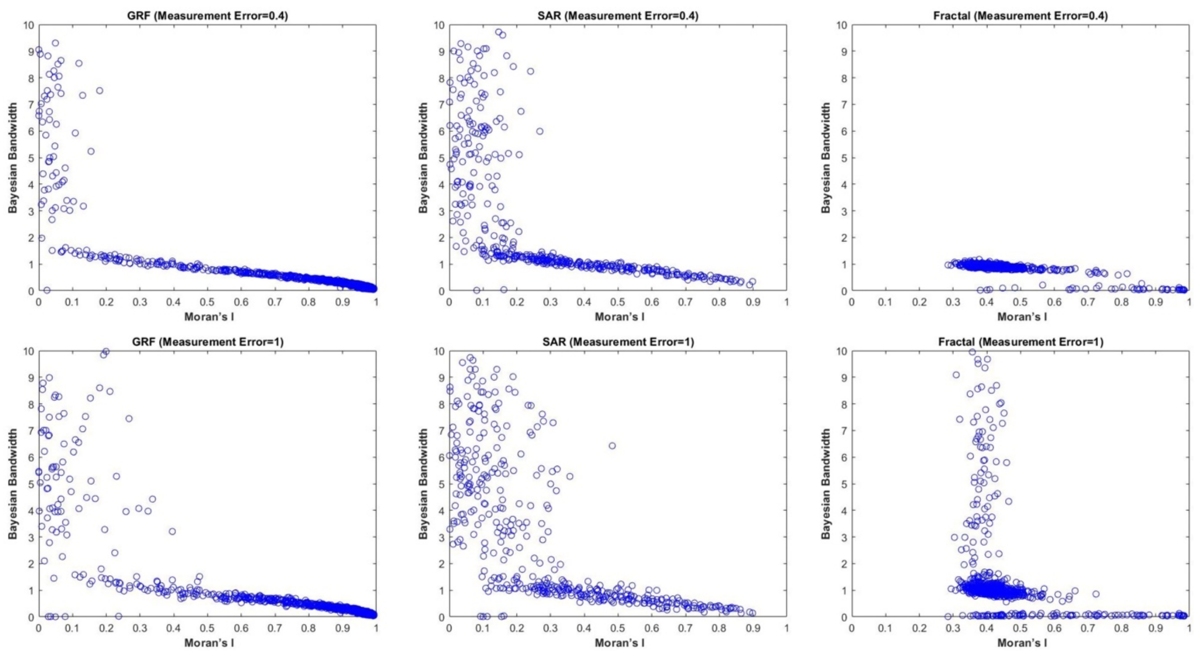

Figure S1: Bandwidth estimates from a SVCM versus Moran's I for three different data generating processes with parameter standard deviation = 1 and two different levels of measurement error (upper row = 0.4; lower row = 1.0).

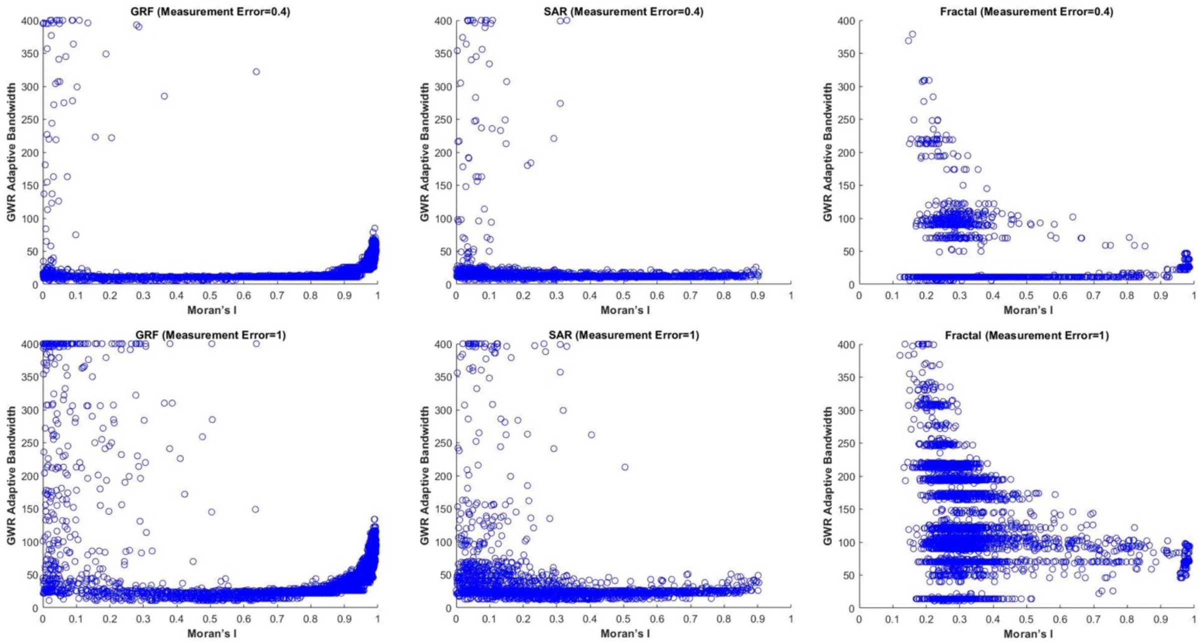

Figure S2: Bandwidth estimates from GWR versus Moran's I for three different data generating processes with parameter standard deviation = 1 and two different levels of measurement error (upper row = 0.4; lower row = 1.0).

Supplemental Material 2: Corroborating the results with a Spatial Gaussian Process model

While the results presented above are reasonable for the MGWR framework, it is useful to compare the behavior of the indicators of scale obtained in an MGWR approach to other approaches. Thus, to corroborate the substantive interpretations we have obtained from MGWR, we estimate a Spatial Gaussian Process model, Gelfand (2003)'s Bayesian Spatially-Varying Coefficient model, in the same configurations as the MGWR models shown above. This is not to exhaustively compare sets of models that purport to provide the same output, but rather to demonstrate that the inferences about the behavior of bandwidth parameters that we make above are likely common to other modelling frameworks.

Further, the calibration of Bayesian spatially-varying coefficient models is more complex than for the equivalent GWR or MGWR models. Though software exists for the calibration of SVC models, it is not as computationally scalable as that for MGWR (Li & Fotheringham, 2020). Consequently, the following results are an abbreviated version of those presented above for the GWR framework. In particular, it is not feasible to reproduce the two-dimensional 'heat maps' (Figures 3, 5 , 6, and 7) of optimal GWR bandwidths for the SVC model. Instead, one-dimensional subsets or "slices" of results akin to Figure 4 are produced to explore the sensitivity of the Bayesian bandwidth by plotting a single system characteristic while holding another constant.

For example, Figures S3 & S4 present the "slice" analogous to the above GWR results (Figures 3, 5 and 6) with standard normal measurement error[6] by plotting the Bayesian (Figs S3, S4) or MGWR (FS S4, S3) bandwidth estimates as a function of Moran's *I* (i.e., spatial dependence) for each of the three DGPs described above for two levels of process variation (0.4 and 1.0). Both indicate that as process dependence increases, the bandwidth estimate decreases, although this is only clearly seen in the SVC case when the process variability is relatively high (bottom row). When Moran's *I* is low, the bandwidth estimate from the SVC model also tends to be relatively invariant to process spatial dependence: in these cases, the process is obscured by large noise or weak parameter variation. An identical presentation for model error is presented in Supplementary Material, and the similarity between the two models holds there as well. Altogether, the similarity across different dimensions of process and parameter variability corroborates our interpretation of the MGWR results presented earlier.

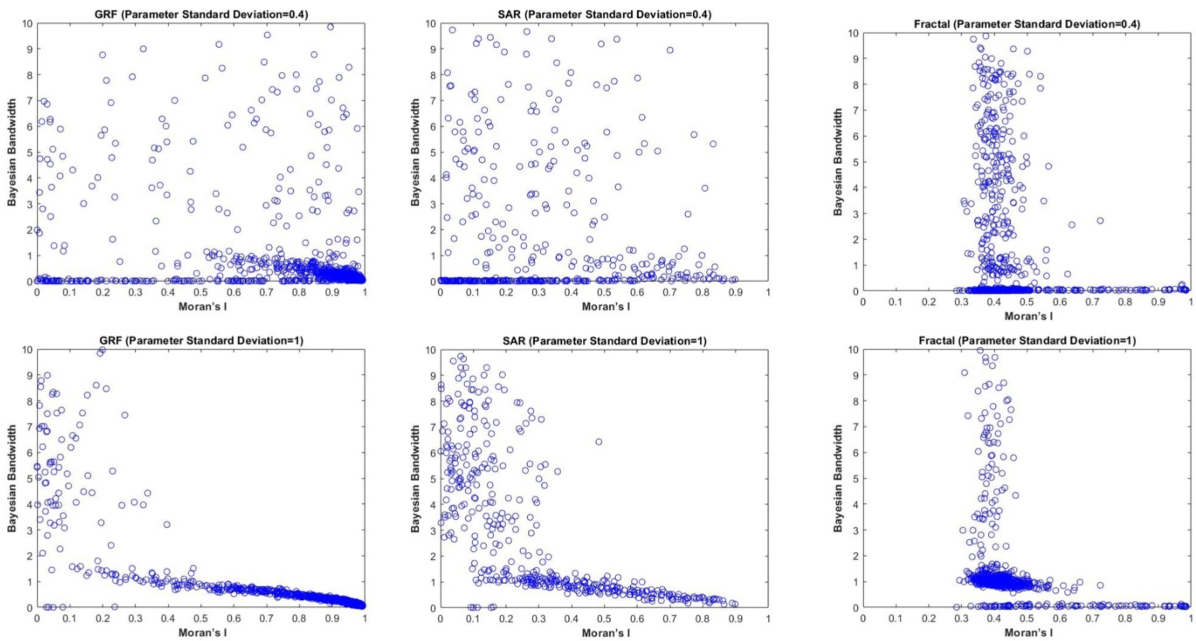

Figure S3: Bandwidth estimates from a SVCM versus Moran's I for three different data generating processes and two different parameter standard deviations (upper row = 0.4; lower row = 1.0) with a standard normal error.

[6] Estimates were made using the spBayes R package (Finley et al. 2015) with priors recommended therein. Overall, 2000 iterations with no thinning were used in the simulations. Due to the generally well-conditioned nature of the study design, chains easily converged within the first 500 iterations, and random spot-inspections showed no issues with convergence in moderate or extreme simulations. The final 500 iterations were used for analysis. Despite the extreme computational improvements made by spBayes authors, simulations were still too time-consuming to generate equivalent results of 100 realizations for every combination of error variance and "scale" parameter for the three data generating processes.

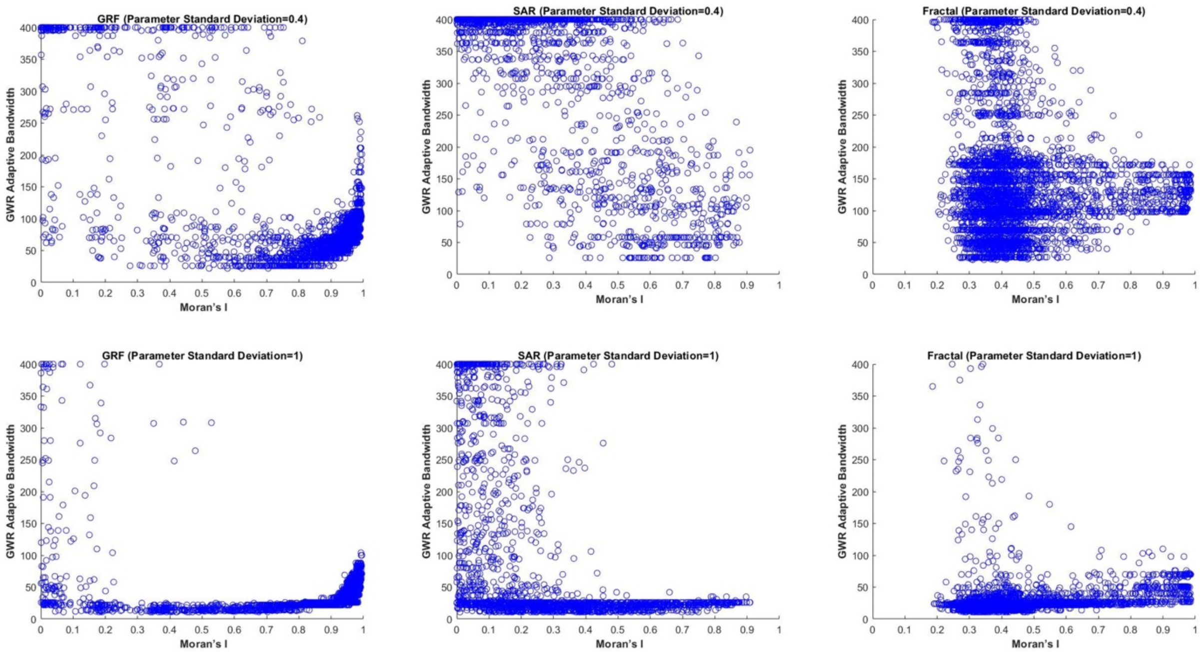

Figure S4: Bandwidth estimates from GWR versus Moran's I for three different data generating processes and two different parameter standard deviations (upper row = 0.4; lower row = 1.0) with a standard normal error.

Supplemental Material 3: Estimated bandwidths with varying degrees of correlation.

We designed an experiment with two covariates with varying degrees of correlation between them to demonstrate the bandwidth sensitivity in a multivariate case. We selected two GRF-based data generating processes, and the parameters of the designed processes and estimated bandwidths are shown in Table S1. We found that the bandwidths remain stable unless the correlation becomes extreme (e.g. R=0.9).

Table S1. Estimated bandwidths with varying degrees of correlation.

GRF	β_1	β_2
Moran's I	0.3	0.7
Parameter standard deviation	0.3	0.7
Correlation coefficient between two covariates	Average bandwidth across 100 realizations	
0.0	330	31
0.1	298	27
0.2	290	35
0.3	312	34
0.4	253	34
0.5	266	40
0.6	289	27

0.7	301	46
0.8	251	66
0.9	210	161