Journal of Geographical Systems
https://doi.org/10.1007/s10109-025-00468-1

ORIGINAL ARTICLE

Scale and correlation in multiscale geographically weighted regression (MGWR)

Wei Kang[1] · **Taylor M. Oshan**[2]

Received: 10 May 2024 / Accepted: 5 May 2025

Abstract
Multiscale geographically weighted regression (MGWR) extends geographically weighted regression (GWR) by allowing process heterogeneity to be modeled at different spatial scales. While MGWR improves parameter estimates compared to GWR, the relationship between spatial scale and correlations within and among covariates—specifically spatial autocorrelation and collinearity—has not been systematically explored. This study investigates these relationships through controlled simulation experiments. Results indicate that spatial autocorrelation and collinearity affect specific model components rather than the entire model. Their impacts are cumulative but remain minimal unless they become very strong. MGWR effectively mitigates local multicollinearity issues by applying varying bandwidths across parameter surfaces. However, high levels of spatial autocorrelation and collinearity can lead to bandwidth underestimation for global processes, potentially producing false local effects. Additionally, strong collinearity may cause bandwidths to be overestimated for some processes, which helps mitigate collinearity but may obscure local effects. These findings suggest that while MGWR offers greater robustness against multicollinearity compared to GWR, bandwidth estimates should be interpreted with caution, as they can be influenced by strong spatial autocorrelation and collinearity. These results have important implications for empirical applications of MGWR.

Keywords Spatial scale · Spatial autocorrelation · Multicollinearity · MGWR · Spatial heterogeneity

JEL Classification C18 · C31

✉ Wei Kang
wei.kang@ucr.edu

[1] Center for Geospatial Sciences, School of Public Policy, University of California Riverside, 900 University Ave, Riverside, CA 92507, USA

[2] Department of Geographical Sciences, University of Maryland, College Park, 2181 LeFrak Hall, College Park, MD 20740, USA

Published online: 06 June 2025

1 Introduction

Spatially heterogeneous processes may vary from location to location in their magnitude and nature. Several strategies have been proposed to capture this spatial nonstationarity using local models because traditional global models are likely to produce misleading results when processes are not stationary. Previous efforts placed these strategies into two categories (Wolf et al. 2018). The first category includes those that typically capture heterogeneity using predefined scales and a relatively small number of discrete areal units or groups at different scales, such as spatial regimes models and hierarchical multilevel models (Banerjee et al. 2014; Anselin 1988). In contrast, the second category includes geographically weighted regression (GWR), eigenvector spatial filter regression, and Bayesian spatially varying coefficients models, which typically capture continuous spatial heterogeneity across a larger number of locations and allow for scales to be selected as part of model fitting (Fotheringham et al. 1998; Gelfand et al. 2003; Griffith et al. 2019). However, it should be noted that this dichotomy is becoming blurred with recent innovations to endogenously select spatially grouped clusters that can express heterogeneity and efforts to combine or blend models from each category (Anselin and Amaral 2023; Sugasawa and Murakami 2021; Hu et al. 2022). This signals a growing acceptance of the notion that spatially heterogeneous processes may come in many different forms and that data-driven techniques can help capture the often multiscale and complex spatial structure that characterizes these processes.

Of the now numerous local modeling methods, GWR and its multiscale extension, multiscale GWR (MGWR), have emerged as one of the most widely used approaches, perhaps due to its relative conceptual and technical simplicity, as well as a suite of accompanying diagnostic tools and accessible software (Fotheringham et al. 2017; Yu et al. 2020a; Fotheringham et al. 2003). Using a spatial kernel function, observations are weighted based on their relative proximity to a set of calibration locations to create local samples and estimate an ensemble of local models. The simplicity of this mechanism has also led to critiques. For example, D. Wheeler and Tiefelsdorf (2005) suggest that GWR is highly susceptible to multicollinearity issues. In response, additional experiments by Páez et al. (2011) and Fotheringham and Oshan (2016) demonstrated that collinearity is less of an issue in GWR when sample sizes are relatively larger. In retrospect, these results are not surprising, and it is clear that small samples (i.e., less than several hundred) are inappropriate for a GWR analysis. Compared to the overall sample, any geographically weighted local sample will necessarily have less variation and smaller effective sample size, a trend that will be exacerbated by the use of increasingly smaller bandwidth values (narrow adaptive kernel). It is easy to quickly run into 'micronumerosity' (Goldberger 1964; 1991). Increasing the overall sample size can increase the initial variation in the data before weighting and may also increase the relative bandwidth size.[1]This generally

[1] This discussion is especially relevant in the context of the adaptive kernel, the most commonly employed kernel in the GWR family. The adaptive kernel is characterized by the number of nearest neighbors utilized for local calibration around each focal observation. Similar principles apply to the fixed kernel scenario, where the bandwidth is determined by the distance to each focal observation.

produces less correlated local samples and reduces the likelihood of selecting very small bandwidths, making each local regression more robust to any potential issues; however, it may not be possible to increase the sample size without extending the study area and reframing the underlying process being modeled, such as when samples are attached to administrative units or grid cells.

Nevertheless, this does not mean that multicollinearity may never be problematic in GWR, and it is certainly possible to find higher or lower levels in the local samples compared to the total sample of observations. Many of the same tools exploited in traditional global regression modeling have been proposed to diagnose and deal with excessive collinearity. Diagnostics such as the condition number, variation inflation factors, correlation coefficients, and variance decomposition proportions have all been extended to examine local collinearity in GWR (Wheeler and Tiefelsdorf 2005; Wheeler 2010). These should be used as part of standard practice, keeping in mind that decisions based on them, such as removing variables, are themselves based on rules-of-thumb rather hard-and-fast cutoffs and each application still requires judgment calls based on acceptable levels of uncertainty, sample size, number of covariates, and prior knowledge of the processes being modeled. Another approach that has been proposed is to extend penalized regression techniques, such as ridge, LASSO, or elastic net to the case of GWR (Wheeler 2009, 2006; Li and Lam 2018). However, these techniques are more complex and computationally intensive when combined with GWR and may have their own challenges for interpreting parameter estimates. In contrast, Fotheringham et al. (2016) suggest using a GWR-specific correction for multiple dependent hypothesis tests and masking cases where the null cannot be rejected to avoid mistakenly interpreting them as part of any potential spatial patterns in the surfaces of parameter estimates. They demonstrate that overall, the risk from collinearity is an elevated number of false positive tests rather than false negative tests.

Another lesson in hindsight is the importance of considering multiple scales. In GWR, each explanatory variable is typically weighted using a kernel function parameterized by the same bandwidth parameter (i.e., the spread of the kernel function). When the bandwidth parameter is selected using a bias-variance trade-off, it can be interpreted as the scale of the modeled process (Fotheringham et al. 2022). Furthermore, the single bandwidth in GWR is selected to optimize this trade-off across all covariates and will tend toward smaller bandwidths when one or more local processes are detected to minimize bias (Yu et al. 2020b). The result is that more smoothing is applied than is necessary for some covariates in the model and this may contribute to the level of collinearity found in local samples. MGWR can alleviate this by allowing a separate bandwidth to be selected for each covariate. Without the need to use one 'compromise' bandwidth, each bandwidth is less restricted in characterizing the spatial scale at which its corresponding process is operating, which means that some bandwidths may be global (i.e., little-to-no

Footnote 1 (continued)
Although increasing the sample size should not affect the bandwidth, it does result in a larger sample size for each local regression.

smoothing), some are regional, and others remain small. In this context, Oshan et al. (2020) report a decrease in the local condition number for MGWR compared to GWR in an empirical application. These results, along with others, are giving way to conventional wisdom regarding the judicious application of GWR, which includes using a global model and then MGWR as a starting point rather than GWR and using local collinearity diagnostics (Fotheringham 2023; Oshan 2023; Fotheringham et al. 2023). Nevertheless, it is still not clear how scale, as measured by bandwidth, impacts collinearity and vice versa. While it is understood that multiple scales can reduce local collinearity, it is important to also understand if collinearity can affect bandwidth selection since the latter is a key model output with practical implications for the measurement of process scale.

Additionally, in contrast to classic regression models that ignore the spatial clustering of processes, spatial models such as MGWR anticipate and leverage such structure. In practice, most spatial variables demonstrate some level of positive spatial autocorrelation with similar values clustering in space. This creates the opportunity for local models to obtain local samples that are pockets of very similar values with little variation even though the overall sample might display substantial variations. If two variables are both spatially autocorrelated and globally collinear (i.e., overall pattern), their local samples are more likely to be very similar, leading to exacerbated collinearity in those local samples. Though the role of spatially correlated covariates has been previously investigated in GWR in the context of local parameter estimates (Murakami et al. 2017; Geniaux and Martinetti 2018), it is not understood how this type of correlation may or may not interact with multicollinearity and scale in MGWR.

This paper therefore studies the relationships between (M)GWR model accuracy, spatial scale, and two types of correlation—collinearity between covariates and spatial autocorrelation in covariates—which has yet to be investigated systematically. The results suggest that both types of correlations have a negligible effect on MGWR performance until they become very strong, and their impacts are cumulative. Overall, MGWR is better at alleviating any potential local multicollinearity issues than GWR due to the varying scales estimated for different parameter surfaces. Two additional insights were obtained regarding scale. First, very high levels of spatial autocorrelation and collinearity may potentially contribute toward the underestimation of scale for certain processes, typically those with the lowest level of spatial heterogeneity (including the intercept), potentially falsely identifying local effects. Second, high levels of collinearity may contribute to the misidentification of anticipated scales. In some cases, when the true process is spatially varying, collinearity may lead the bandwidth to be overestimated, which can help mitigate issues associated with high collinearity but may also lead to difficulty identifying local effects. In other cases, when the true process is constant, collinearity may lead the bandwidth to be underestimated but is not problematic for accurately estimating coefficients unless collinearity becomes high.

The remaining sections of the paper are organized as follows. Section 2 introduces the GWR and MGWR models. Section 3 describes the design of the two sets of simulation experiments for a systematic investigation of the interplay between collinearity, spatial autocorrelation in covariates, and spatial scales in MGWR and

GWR. Results from the simulation experiments are given in Sect. 4, and finally, implications of the results and conclusions are discussed in Sect. 5.

2 Local spatial modeling: GWR and MGWR

GWR and MGWR are extensions of the classic multiple linear regression model where the expectation of the outcome or response variable is a linear combination of covariates. These two local spatial models explicitly address process spatial heterogeneity and allow the parameters to be spatially varying by exploiting the spatial dependence structure across observations. They can be formalized as

$$y_i = \sum_{j=0}^{m} \beta_{ij} x_{ij} + \varepsilon_i \tag{1}$$

where a total of m explanatory variables are used in the regression model, y_i is the response variable, and ε_i is the random noise at location i, x_{ij} is the value of the jth explanatory variable at location i and β_{ij} is the associated coefficient (i.e., jth coefficient at location i) free to vary geographically. For GWR, the spatial heterogeneity of the regression coefficients is achieved by calibrating a weighted ordinary least square (OLS) model for each location using spatial weights determined by a spatial kernel and with a bandwidth optimized using cross-validation or the minimization of a model fit statistic (e.g., AICc) during the model calibration process. One of the most widely used kernels, the adaptive bisquare kernel, which is also used here, is defined as

$$w_{ih} \begin{cases} \left[1 - \left[\frac{d_{ih}}{G_i}\right]^2\right]^2, & \text{if} d_{ih} < G_i \\ 0, & \text{Otherwise} \end{cases} \tag{2}$$

This kernel operationalizes the bandwidth (b) as the number of nearest neighbors used for calibration for each location. G_i is the distance from the location i to its bth nearest neighbors and the spatial weight between locations i and h is negatively associated with the distance between these two locations. In GWR, the same bandwidth (b) is used for constructing the spatial weights for all the covariates, therefore assuming that all the modeled relationships operate at the same spatial scale.

MGWR relaxes the assumption of a single scale for all the modeled processes and allows each coefficient β_j to have a separate bandwidth (i.e., $b_0, b_1, b_2, \ldots, b_m$). The bandwidths are estimated in the model calibration process and are interpreted as the spatial scale at which the conditional relationship (or spatial process) operates. The smaller the bandwidth, the more localized the spatial process and vice versa. MGWR model estimation is operationalized using a generalized additive model (GAM) and a backfitting algorithm to calibrate each relationship iteratively (Fotheringham et al. 2017). Inference about each surface of local parameter estimates can then be adjusted for multiple dependent hypothesis tests based on the associated bandwidth (Yu et al. 2020a) in order to more flexibly discern patterns from noise in the surfaces of local parameter estimates. It is also possible to conduct Monte Carlo tests on the significance of the spatial variability of each surface (Fotheringham et al. 2002) and construct 95% confidence intervals on the bandwidth estimates (Li et al. 2020).

3 Experimental design

3.1 Data generating process

The data generating process (DGP) utilized followed an MGWR-like regression specification that incorporates two covariates (X_1 and X_2), one intercept parameter (β_0), and two slope parameters (β_1 and β_2). The three parameters are configured to potentially include varying levels of spatial heterogeneity. The DGP is formally defined as

$$y_i = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \varepsilon_i \tag{3}$$

where the error term follows a normal distribution centered at 0 $\left(\varepsilon \sim N\left(0, \sigma^2 I\right)\right)$.

The two covariates are spatially configured to exhibit varying levels of spatial autocorrelation and varying degrees of collinearity with one another. After describing two sets of known parameter surfaces, methods for generating collinear and spatially autocorrelated covariates are outlined.

3.2 True parameter surfaces

Two sets of parameter surfaces were employed, which differ in sample size and degree of spatial heterogeneity. The first is identical to the true surfaces used in the initial MGWR simulation experiments in Fotheringham et al. (2017) with the first two surfaces swapped to make sure second and third parameter surfaces exhibit a greater difference in the level of spatial heterogeneity (top row of Fig. 1). This adjustment allows the intercept to be more thoroughly investigated, which is important due to the unique role it plays in a MGWR model. These surfaces are configured in a (25, 25) regular lattice and range in values from 1 to 5. Though the second surface is constant across space and represents a global process, the other two surfaces vary across space at different scales, with β_0 demonstrating a lower level of spatial heterogeneity (the rate of change over space is lower) and β_2 demonstrating a higher level of spatial heterogeneity (the rate of change over space is higher). Additionally, the original scenario was considered in which the intercept surface is constant following Fotheringham et al. (2017) (Fig. S1), which serves as a point of comparison for the first simulation experiment.

The second set of parameter surfaces follows the design of Fotheringham and Oshan (2016) which was used to previously study multicollinearity and GWR. As shown in the bottom row of Fig. 1, these surfaces have a larger sample size and are based on a (50, 50) regular lattice. Though the intercept surface

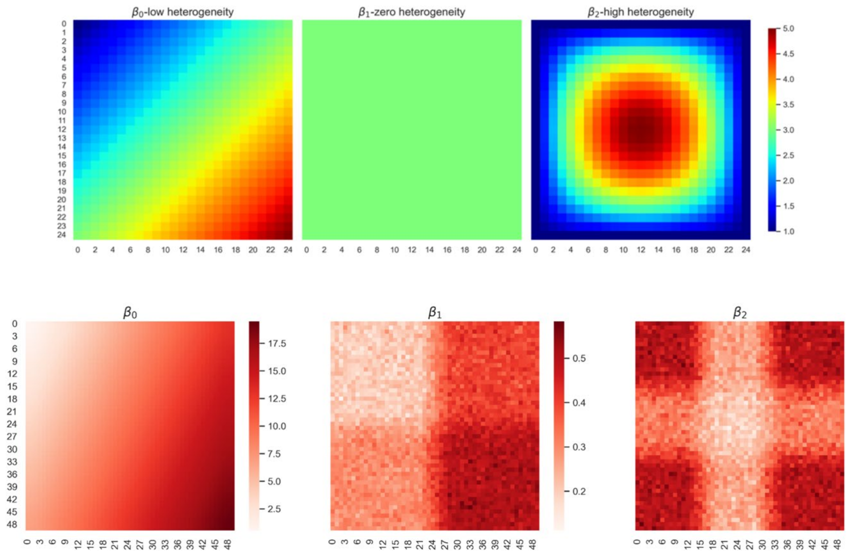


Fig. 1 Two different sets of true parameter surfaces for the simulation experiment-based varying patterns and sample sizes

possesses the lowest level of perceived spatial pattern heterogeneity, it actually has a much wider overall range of values, producing the highest relative level of spatial heterogeneity among the three surfaces.

3.3 Spatially autocorrelated covariates

To simulate realistic covariates that follow Tobler's first law of geography, "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970), a first-order spatial autoregression specification was used and is defined as

$$X = \rho WX + \epsilon \tag{4}$$

which can be rewritten in reduced form as

$$X = (I - \rho W)^{-1}\epsilon \tag{5}$$

where $\rho \in (-1, 1)$ is the spatial autoregressive parameter representing the strength of spatial dependence, ϵ is a white noise vector $\left(\varepsilon \sim N\left(0, \sigma_1^2 I\right)\right)$ assumed to follow a normal distribution, I is a (n, n) identity matrix with 1's on the diagonal and 0's elsewhere, and W is the (n, n) row-stochastic spatial weight matrix expressing interactions across spatial units. Queen contiguity was used for W such that any polygons that share a common edge or vertex are considered neighbors.

3.4 Two types of correlation: spatial autocorrelation and collinearity

The two covariates used in the DGP are constructed to potentially have both spatial autocorrelation and collinearity. After simulating the first explanatory variable (X_1) using the spatial autoregressive model in Eq. (5), the same formula is also used to generate an intermediate variable Z. Then, the second explanatory variable (X_2) is simulated using Z so that both explanatory variables will be approximately correlated with a set level denoted by a given Pearson's r value, as well as spatially autocorrelated at the same level (ρ). The complete process is formulated as

$$\begin{aligned} X_1 &= (I - \rho W)^{-1}\epsilon_1 \\ Z &= (I - \rho W)^{-1}\epsilon_2 \\ X_2 &= rX_1 + \sqrt{1 - r^2}Z \end{aligned} \tag{6}$$

The two error terms ϵ_1 and ϵ_2 are independent and each follows a standard normal distribution ($N(0,1)$). Values of ρ for spatial dependence and r for bivariate correlation were both allowed to take on values of {0, 0.5, 0.8, 0.85, 0.9, 0.95}.[2] It should be noted that the spatial autoregressive coefficient ρ used in the DGP is not equivalent to Moran's I statistic though they are positively associated.[3] The restriction of a consistent spatial autocorrelation level for the two covariates can be relaxed in the scenario where there is a lack of global collinearity between the two covariates, resulting in different values of ρ for X_1 and X_2. This design helps disentangle the sole impact of spatial autocorrelation.

The generated covariates (X_1 and X_2) were standardized before being used within the DGP in Eq. (3) to make sure the optimal bandwidths estimated from MGWR reflect only the spatial scale at which each process is operating. For each combination of parameters, 100 simulations were generated using a normally distributed random error with a mean of 0 and a standard deviation (σ) of 1.5 for the smaller sample and 0.5 for the larger sample scenarios. This gives rise to two different average levels of signal-to-noise ratio (SNR) resembling a pseudo-R^2 (defined as 1 minus the ratio of noise variance (σ^2) to signal variance ($\mathrm{Var}(X\beta)$), which are 0.87 and 0.221, respectively. An MGWR model and a GWR model were calibrated for each simulated realization. For the smaller sample size case, an additional simulation experiment was carried out with a larger standard deviation of the error term (i.e., 2.5) to incorporate a higher level of random noise and thus a smaller level of SNR

[2] Smaller values, such as 0.1, 0.2, 0.3, and 0.4, were tested for both the spatial autoregressive parameter and the bivariate correlation. Since these small values had a negligible effect on the performance of GWR and MGWR, only results for larger values were reported. Additionally, an extreme value of 0.99 was included in the initial experiments. However, these cases resulted in highly pathological scenarios unsuitable for any realistic application and indeed lead to numerous local optima, making convergence of the MGWR calibration algorithm unlikely. As a result, this value was excluded from the paper.

[3] While the spatial autoregressive term ρ in the DGP and Moran's I on the generated spatial patterns is positively associated (Pearson's correlation coefficient = 0.93), the latter is smaller (Moran's I is about 67% of ρ). As shown in Fig. S2, 0.8 is correspondent to a 0.5 estimate of Moran's I for the scenario of (25,25) regular lattice and queen contiguity weights.

(i.e., 0.636). Model calibration and inference were carried out using the open-source Python package—PySAL/mgwr version 2.2.1 (Oshan et al. 2019).

3.5 Evaluation criteria

One of the primary objectives of this study is to understand the link between the estimation of spatial process scale and the two types of correlations (spatial autocorrelation and bivariate collinearity). Therefore, an examination of the optimal bandwidths estimated from MGWR (i.e., b_0, b_1, and b_2) and GWR (i.e., b) will first be conducted and compared with the levels of spatial heterogeneity observed in the true parameter surfaces.

The ability of GWR and MGWR to recover the true parameter surfaces is measured with two relative metrics. The first metric relies on the root-mean-squared error (RMSE), which compare the estimates with the known parameters for each realization. For each parameter (β_j) and realization (s), $RMSE_{sj}$ is calculated as follows:

$$\mathrm{RMSE}_{sj} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\beta_{ij} - \hat{\beta}_{sij}\right)^2} \tag{7}$$

where n is the number of observations and $\hat{\beta}_{sij}$ is the estimate of parameter β_j for the ith observation in the sth realization. The relative metric rRMSE is defined in Eq. (8) as the ratio of the average $RMSE$ for the scenario under investigation (e.g., $\rho = 0.9,\ r = 0.8$) to the average RMSE obtained for the reference scenario (i.e., $\rho = 0,\ r = 0$) for MGWR and GWR, respectively. This metric yields a value of 1 for the reference scenario and is expected to increase as either correlation grows. Assessing how well GWR and MGWR can produce lower magnitudes of this metric provides insight into their ability to address increasing correlation.

$$\mathrm{rRMSE} = \frac{\frac{1}{3}\frac{1}{100}\sum_{j=0}^{2}\sum_{s=1}^{100}\mathrm{RMSE}_{sj}}{\frac{1}{3}\frac{1}{100}\sum_{j=0}^{2}\sum_{s=1}^{100}\mathrm{RMSE}_{sj}(\rho = 0, r = 0)} \tag{8}$$

The other relative metric, the mean absolute percentage error (MAPE), measures the accuracy of a model's parameter estimations in percentage terms and is defined in Eq. (9):

$$\mathrm{MAPE}_{sj} = 100\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\beta_{ij} - \hat{\beta}_{sij}}{\beta_{ij}}\right| \tag{9}$$

This metric ranges from 0 to 100, and a smaller value indicates a more accurate prediction/recovery of the parameter surface and vice versa.

A local multicollinearity diagnostic, namely, the local condition number, is used to detect potential local multicollinearity issues in GWR and MGWR (D. C.

Wheeler 2006)[4]. This diagnostic relies on the singular-value decomposition (SVD) of the design matrix X for each location i. For each location i, the SVD is defined as follows

$$W_{(i)}^{\frac{1}{2}}X = \mathrm{UDV}^{T} \tag{10}$$

where U and V are orthogonal $(n, m+1)$ and $(m+1, m+1)$ matrices, D is a $(m+1, m+1)$ matrix with a diagonal of decreasing singular values, $W_{(i)}^{\frac{1}{2}}$ is the square root of the diagonal spatial weight matrix for location i obtained from the adaptive bisquare kernel function in Eq. (2). The local condition index (LCI) for the jth explanatory variable at location i is defined as

$$\mathrm{LCI}_{ij} = \frac{d_{\max(i)}}{d_{(i)j}} \tag{11}$$

where $d_{(i)j}$ is the j th singular value in D and $d_{\max(i)}$ is the largest singular value. The largest *LCI* for this location is defined as the local condition number (LCN_i). The larger the LCN, the stronger the local collinearity among the covariates and the intercept. A critical value of 30 is often used to determine major issues, while 10 is a more conservative threshold to suggest little-to-no collinearity issues. Thresholds of 10, 20, and 30 are considered here.

4 Results

4.1 Simulation experiment one

In the first simulation experiment, which has a smaller sample size ($n = 625$), the true surface associated with the first covariate is spatially homogeneous, and the intercept and the other true parameter surface have a similar numerical scale but different levels of spatial heterogeneity. When the covariates are neither spatially autocorrelated nor collinear, the estimated optimal bandwidths for all three parameter surfaces correctly reflect the different levels of spatial heterogeneity in the true parameter surfaces, with the bandwidth estimated for the second parameter surface tending toward the maximum global value and the other two being much more local. The absolute magnitudes of the spatial scales, as measured by the optimal number of nearest neighbors used for local model calibration, increased with stronger noise (Table 1). On the other hand, both types of association tend to have little-to-no impact on the performance of MGWR or GWR until either becomes very strong (Figs. 2 and 3), at which point some additional patterns emerge.

[4] The local variance inflation factor (VIF) is another multicollinearity diagnostics available for GWR but not MGWR and cannot be used for assessing local collinearity with the intercept term.

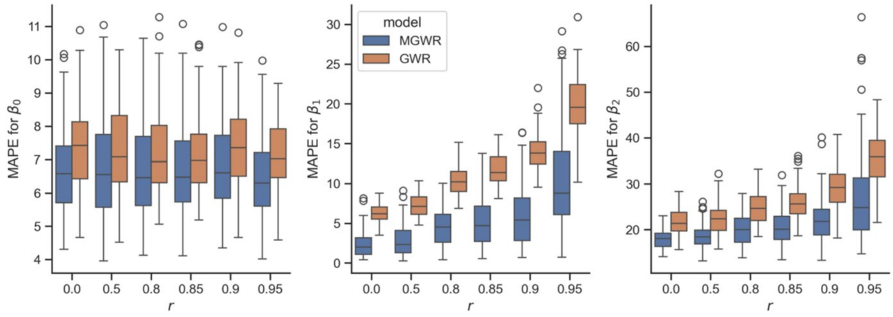


Fig. 2 MAPEs for GWR and MGWR when both covariates are exempt from spatial autocorrelation ($n = 625, \sigma = 1.5$)

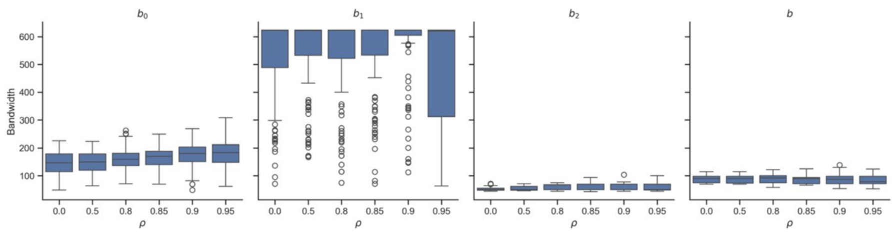


Fig. 3 Optimal bandwidths for GWR (b) and MGWR (b_0, b_1, b_2) when two covariates are spatially autocorrelated but are not linearly associated ($n = 625, \sigma = 1.5$)

Table 1 Median optimal bandwidth estimates from MGWR and GWR for different SNR levels ($n = 625, \rho = 0, r = 0$)

Constant parameter surface	SNR	σ	MGWR bandwidth estimates			GWR bandwidth estimate
			b_0	b_1	b_2	b
β_0	0.874	1.5	623	148	50	90
β_0	0.655	2.5	624	227	71	139
β_1	0.87	1.5	147	623	50	90
β_1	0.636	2.5	197	623	73	130

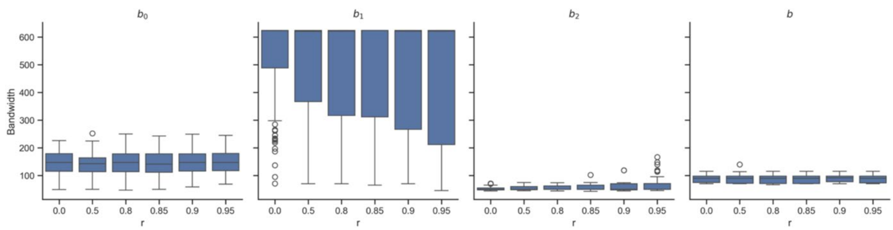


Fig. 4 Optimal bandwidths for GWR (b) and MGWR (b_0, b_1, b_2) when both covariates are exempt from spatial autocorrelation ($n = 625, \sigma = 1.5$)

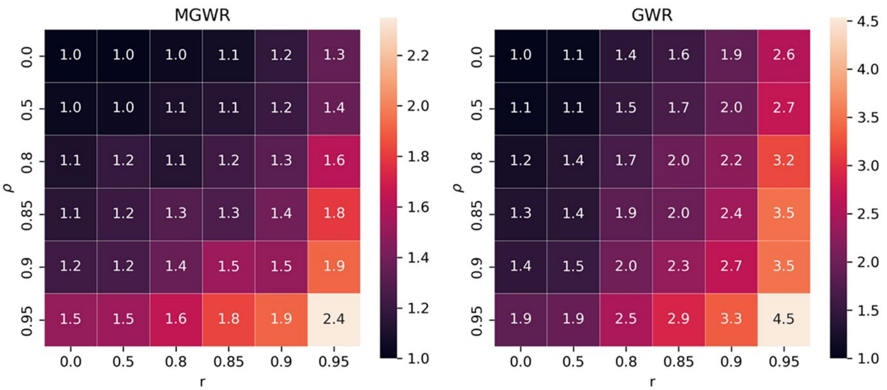
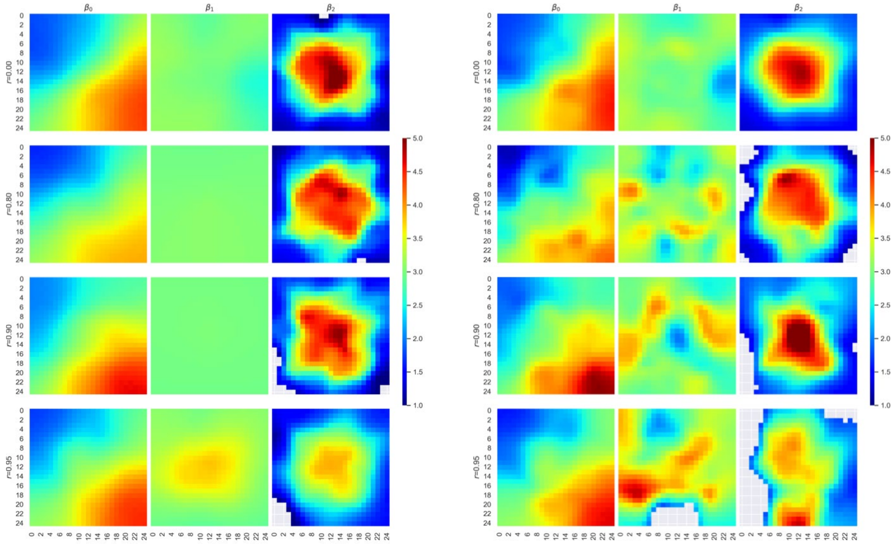


Fig. 5 rRMSEs for GWR and MGWR ($n = 625, \sigma = 1.5$)



(a) MGWR

(a) GWR

Fig. 6 An example of parameter estimates for scenarios where both covariates are exempt from spatial autocorrelation ($n = 625, \sigma = 1.5$) (statistically insignificant estimates filtered)

4.1.1 Impact of collinearity

To isolate the potential impact of collinearity on (M)GWR, the DGP was controlled so that the covariates were not spatially autocorrelated. Collinearity has essentially no impact on the estimation of the parameter surface that is uncorrelated (i.e., intercept) and its spatial scale (Fig. 2 and 4). In contrast, for the two collinear covariates, while the bandwidth estimates for the associated parameters are affected to a

Table 2 Average percentage of locations that have a local condition number > 10 using GWR ($n = 625$)

		Collinearity (r)					
		0.00	0.50	0.80	0.85	0.90	0.95
Spatial autoregressive parameter (ρ)	0.00	0.0	0.0	0.0	0.0	0.0	0.98
	0.50	0.0	0.0	0.0	0.0	0.01	2.6
	0.80	0.0	0.0	0.06	0.36	1.97	17.73
	0.85	0.0	0.0	0.23	1.06	4.49	25.05
	0.90	0.0	0.01	1.29	3.01	9.28	32.19
	0.95	0.43	1.02	8.41	14.11	24.72	51.33

minimal degree (Fig. 4), the parameters for these two covariates have increasing estimation error as collinearity increases and the impact is stronger for GWR than MGWR (Fig. 2 and 5). However, this does not lead to a poorer recovery of the true parameter surfaces except in the most extreme scenarios (i.e., $r = 0.95$) and even then, the distortion is moderate for MGWR (Fig. 6)[5]. Avoiding such high levels of collinearity is a nearly universal guideline for any regression exercise, and these results reinforce previous results that that collinearity alone is not inherently problematic (Fotheringham and Oshan 2016).

4.1.2 Impact of spatial autocorrelation

In contrast, when there is no collinearity between the two covariates, spatial autocorrelation in covariates has spillover effects on the estimation accuracy of the intercept and the associated bandwidth. There are increased levels of estimation error for all three parameters as the spatial autocorrelation increases (Fig. 7). Although the bandwidth estimates for the intercept slightly increase as the spatial autocorrelation grows, the bandwidth estimates for the other two parameters seem to be robust to the increasing spatial autocorrelation (Fig. 3). Despite the increased estimation error, MGWR still produces slope parameter surfaces that preserve their distinct spatial patterns under high levels of spatial autocorrelation (Fig. S3). The scenario where one covariate was controlled to be spatially random, while the second covariate was controlled to be spatially autocorrelated was also explored, which produced similar results in regard to the intercept and autocorrelated covariate, while the parameter estimates and associated bandwidth for the random covariate remained essentially unaffected. Similar patterns were observed for the experiments where the true intercept is spatially constant; the bandwidth estimation is robust to the level of spatial autocorrelation (Fig. S4).

[5] For all results, statistically insignificant estimates at the 5% significance level are not displayed and multiple testing is adjusted for using methods from Yu, Fotheringham, Li, Oshan, Kang, et al. (2020).

4.1.3 Spatial autocorrelation and collinearity

The combined effects of both types of correlation have never previously been examined for (M)GWR, and there appears to be a cumulative impact on the results compared to each individual factor. As both types of correlation get stronger, it takes many more iterations for MGWR to converge. Tables 2 and 3 describe the average percentage of local condition numbers (LCNs) greater than 10 across simulations

Table 3 Average percentage of locations that have a local condition number > 10 using MGWR ($n = 625$)

		Collinearity (r)					
		0.00	0.50	0.80	0.85	0.90	0.95
Spatial autoregressive parameter (ρ)	0.00	0.0	0.0	0.0	0.0	0.0	0.06
	0.50	0.0	0.0	0.0	0.0	0.0	0.08
	0.80	0.0	0.0	0.01	0.01	0.06	0.57
	0.85	0.0	0.0	0.0	0.02	0.16	1.77
	0.90	0.0	0.0	0.04	0.2	0.63	2.82
	0.95	0.01	0.01	0.63	0.61	1.8	4.69

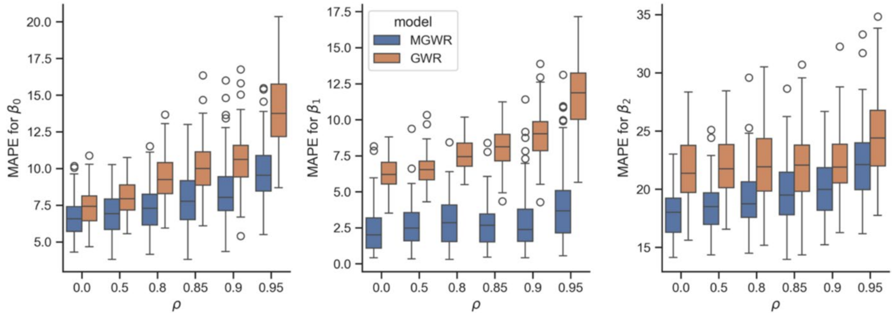


Fig. 7 MAPEs for GWR and MGWR when two covariates are spatially autocorrelated but are not linearly associated ($n = 625$, $\sigma = 1.5$)

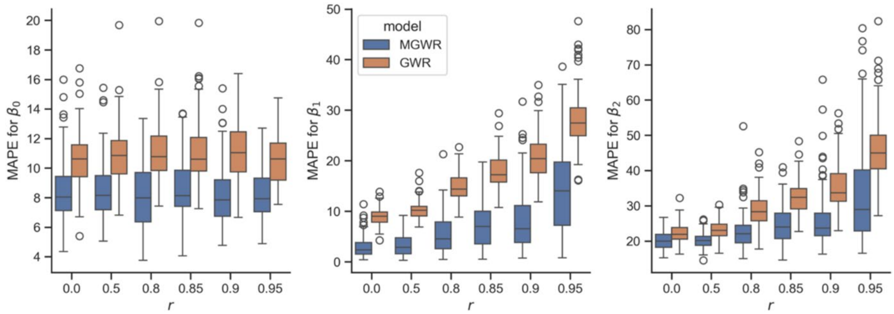


Fig. 8 MAPEs for GWR and MGWR for increasing collinearity when both covariates are spatially autocorrelated ($\rho = 0.9$, $n = 625$, $\sigma = 1.5$)

for GWR and MGWR, respectively. In general, as either collinearity is increased (across columns) or autocorrelation is increased (across rows), the occurrence of local collinearity increases and is most pronounced when both types of correlation are high. However, the overall presence of local collinearity is always lesser for MGWR than for GWR for all scenarios and this is likely due to different bandwidths for different covariates leading to lower levels in some locations (Fig. 7).

As shown in Fig. 8 and 9, when either (a) the spatial autocorrelation level in both covariates is high ($\rho = 0.9$) and collinearity is increased or (b) the collinearity is high ($r = 0.9$) and the spatial autocorrelation level in both covariates is increased, there are higher levels of estimation error for the parameters compared to Figs. 2 and 3 when there are only increased levels of one or the other. As both types of correlation get very strong (e.g., $r > 0.9$ and $\rho > 0.9$), the two covariates exhibit very similar spatial patterns, both globally and locally. This similarity could complicate parameter estimation, leading to instances where spatial patterns of parameter surfaces with higher spatial heterogeneity are absorbed by those with lower heterogeneity (Fig. 10). This effect is further indicated by the significantly smaller bandwidth estimated for the constant surface of β_1 (Fig. S5). Therefore, for this smaller sample experimental design, there seems to be a link between joint levels of high collinearity and autocorrelation, increased local collinearity, and increased estimation error.

4.1.4 Robustness check

Several additional simulation experiments were conducted as a robustness check on the patterns observed with two explanatory variables. In the first additional experiment, a third explanatory variable (X_3) was introduced, generated independent of the other two correlated and spatially autocorrelated variables (X_1 and X_2). The associated parameter surface (β_3) exhibits a similar level of spatial heterogeneity to the intercept, and all the four "true" parameter surfaces are uncorrelated (Fig. 11). The experiment aimed to investigate whether the biases observed in the coefficient estimates of the two spatially correlated and/or collinear covariates affect the estimation of an independent variable. The results indicate that spatial autocorrelation (ρ) in the other two explanatory variables (β_1 and β_2) and

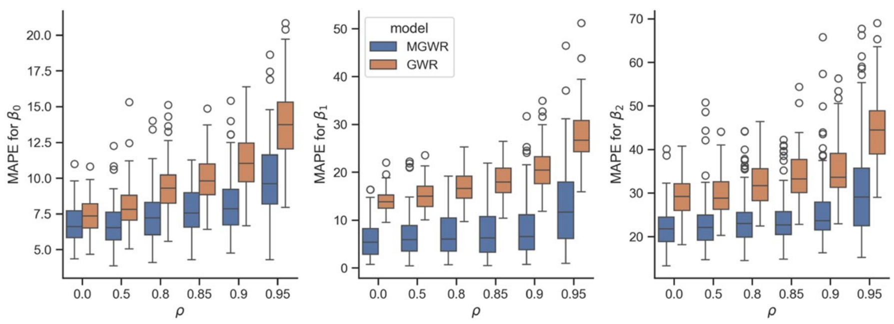


Fig. 9 MAPEs for GWR and MGWR for increasing spatial autocorrelation with fixed collinearity ($r = 0.9$, $n = 625$, $\sigma = 1.5$)

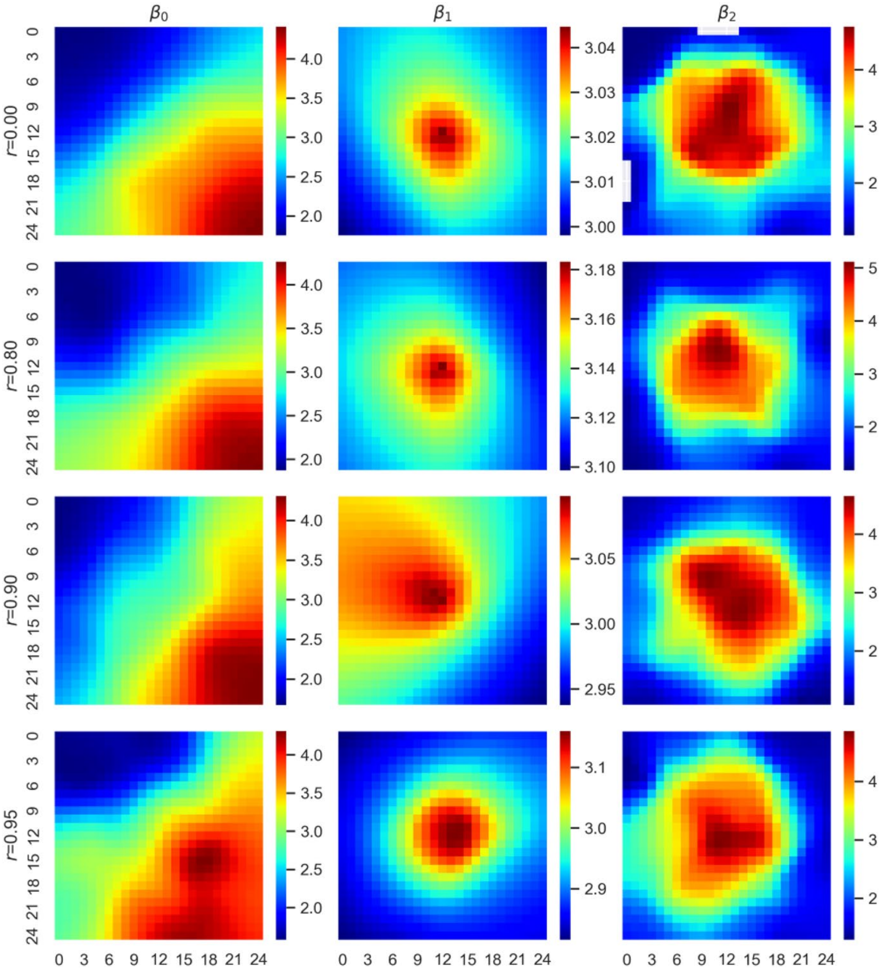


Fig. 10 An example of MGWR parameter estimates for scenarios where both covariates are highly spatially autocorrelated ($n = 625, \sigma = 1.5, \rho = 0.9$) (statistically insignificant estimates filtered)

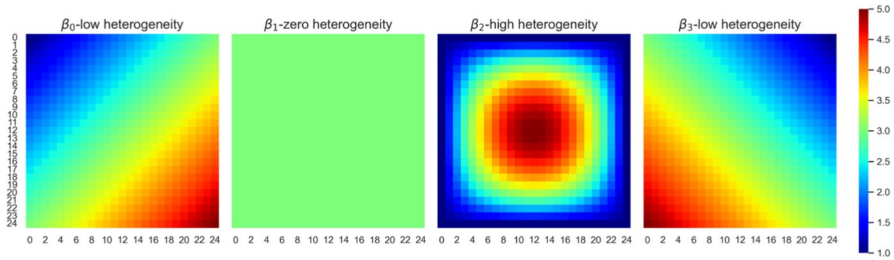


Fig. 11 True parameter surfaces for the first additional simulation experiment

their bivariate correlation (r) do not impact the bandwidth selection or parameter estimation for the third explanatory variable (β_3).

The second additional experiment follows the simulation experiment one in terms of the DGP for the covariates. However, unlike the original design, where the 'true' parameter surfaces associated with the two covariates were independent, this experiment introduces a high level of correlation between β_1 and β_2. Specifically, these two parameter surfaces were using the DPG in Eq. (6), where β_1 and β_2 are both spatially autocorrelated to a similar extent ($\rho_1 = \rho_2 = 0.95$), and they are positively associated ($r_{12} = 0.8$) (Fig. 12). The estimated Pearson's correlation coefficients among the "true" β s are $Corr\left(\beta_0, \beta_1\right) = -0.34$, $Corr\left(\beta_0, \beta_2\right) = -0.28$, and $Corr\left(\beta_1, \beta_2\right) = 0.7$. The DGP of the explanatory variables (X_1 and X_2) follows the structure of the original experiments, meaning that X_1 and X_2 could be both spatially autocorrelated and linearly associated, with $\rho, r \in \{0,0.5,0.8,0.85,0.9,0.95\}$.

The overall conclusion regarding the impact of the two types of correlation remains valid. Specifically, spatial autocorrelation in covariates increases estimation errors for both the associated parameters and the intercept, whereas collinearity increases estimation errors only for the associated parameters. When collinearity and spatial autocorrelation in covariates are absent, both GWR and MGWR successfully preserve the 'true' correlation among parameter surfaces. However, as these correlations become stronger, MGWR is better positioned to generate more accurate local estimates and maintain the correlation between parameter surfaces (Fig. S6).

4.2 Simulation experiment two

For the second experiment, the sample size is larger ($n = 2500$) and none of the true surfaces is constant. The intercept exhibits distinct spatial patterns (gradually changes from the smallest value in the upper-left corner to the largest value in the lower-right corner) and possesses a larger numerical scale than the other two true parameter surfaces.

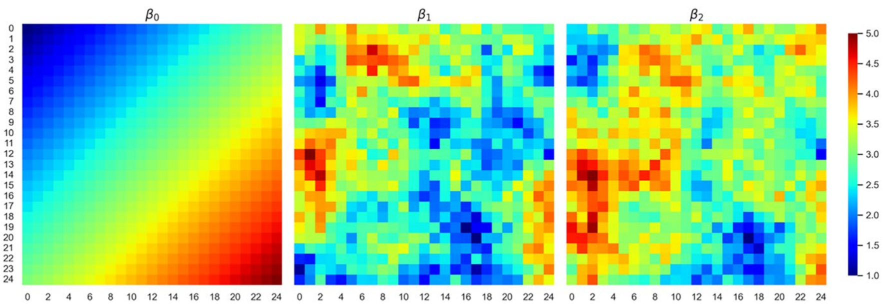


Fig. 12 True parameter surfaces for the second additional simulation experiment

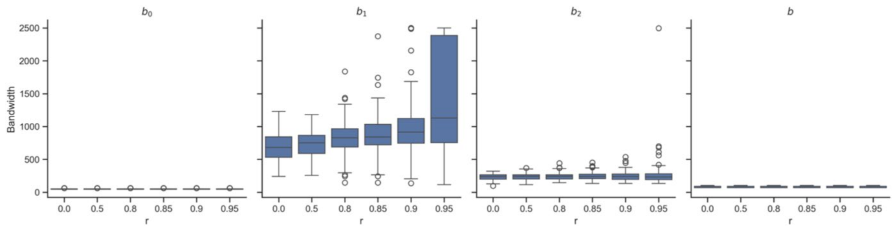


Fig. 13 Optimal bandwidths for GWR (b) and MGWR (b_0, b_1, b_2) when both covariates are exempt from spatial autocorrelation ($n = 2500$)

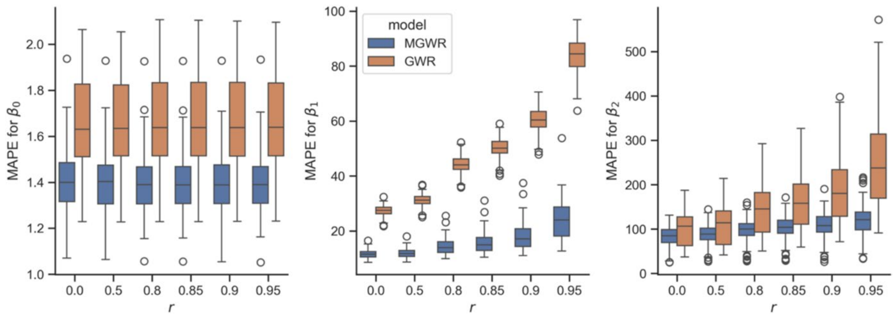
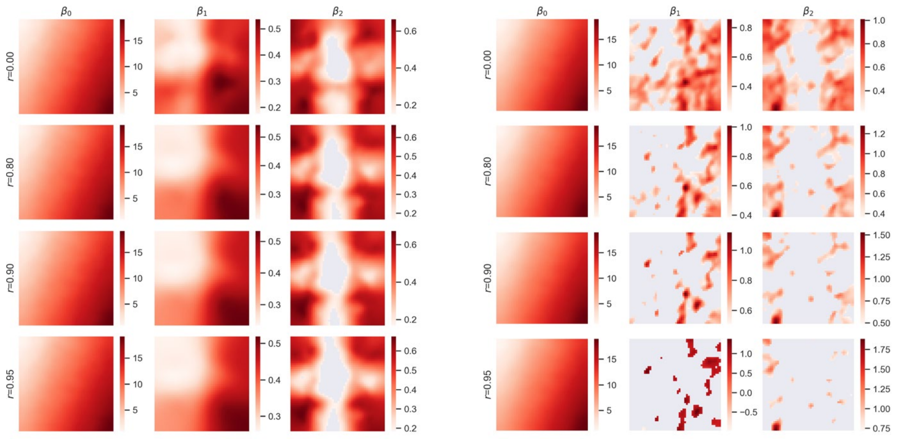


Fig. 14 MAPEs for GWR and MGWR when both covariates are exempt from spatial autocorrelation ($n = 2500$)

4.2.1 Impact of collinearity

Similar to the trend from the smaller sample experiment, when the two covariates are randomly spatially distributed, increasing collinearity increases estimation error for coefficients of the two covariates without much effect on the intercept and the increase is much less pronounced for MGWR than for GWR; however, there are some distinct differences regarding the estimated bandwidths (Fig. 13 and 14). Notably, the bandwidth estimates for coefficients of the two collinear covariates that exhibit a lower level of spatial heterogeneity are inflated as the collinearity increases and especially as it becomes very strong (e.g., $r > 0.9$), which is also observed in the additional smaller sample experiment where the intercept is spatially homogeneous (Fig. S7). It suggests that MGWR may attempt to compensate for the shrunken effective sample size due to correlation by increasing the number of samples for each local estimation through the use of a larger bandwidth. This could also be explained in terms of a trade-off between either information and misinformation or bias and variance during bandwidth selection (Fotheringham et al. 2022). High collinearity suggests that more local samples may contain less information, while more global samples may contain less misinformation. This decreases the bias reduction and increases the variance inflation associated with smaller bandwidths (i.e., reduces the benefits and increases penalty) while potentially also introducing less bias but reducing some variance (i.e., reduces the penalty and increases



(a) MGWR (b) GWR

Fig. 15 An example of parameter estimates for scenarios where both covariates are exempt from spatial autocorrelation ($n = 2500$) (statistically insignificant estimates filtered)

Table 4 Average percentage of locations that have a local condition number > 10 using GWR ($n = 2500$)

		Collinearity (r)					
		0.00	0.50	0.80	0.85	0.90	0.95
Spatial autoregressive parameter (ρ)	0.00	0.0	0.0	0.0	0.0	0.0	1.12
	0.50	0.0	0.0	0.0	0.0	0.05	4.27
	0.80	0.0	0.0	0.15	0.63	3.09	21.15
	0.85	0.0	0.01	0.75	2.2	7.22	30.24
	0.90	0.04	0.16	3.55	7.21	16.08	42.41
	0.95	1.81	3.45	15.02	21.94	33.85	58.98

benefits) when incorporating samples from further away. The outcome is that with increasing collinearity (i.e., reduction in overall information), it becomes more important to reduce variance than decrease bias and this is achieved by selecting a larger bandwidth and increasing the sample size. As a result, MGWR can recover some of the patterns in the true surfaces even for the most pathological case (e.g., $r = 0.95$) (Fig. 15[6]), although the estimates for β_1 have little-to-no overall variation compared to the true surface and are essentially an average. The trade-off is that in return MGWR is much less prone than GWR to potential local collinearity issues

[6] The missing areas in Fig. 14 are statistically insignificant parameter estimates that have been filtered and correspond to small true values with low signals.

(Tables 3–4). In fact, for MGWR, collinearity has a negligible impact on potential local collinearity issues except in the most extreme case (e.g., $r = 0.95$).

4.2.2 Impact of spatial autocorrelation

The general impact due solely to spatial autocorrelation in both covariates is similar for both MGWR and GWR in that it still increases estimation error for all three surfaces, but there is no longer an issue with misestimation of the intercept bandwidth, even for the most extreme scenario ($\rho = 0.95$) (see Fig. S8, S9, and S10). Instead, there is some over- or under-estimation of the bandwidth for β_1 and β_2 similar to the impact of collinearity but much more moderate. For both GWR and MGWR, there are virtually no potential local collinearity issues (due solely to spatial autocorrelation or global collinearity) except for GWR for the most extreme scenario.

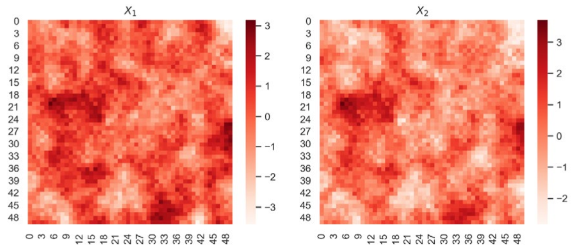
4.2.3 Spatial autocorrelation and collinearity

The joint impact of spatial autocorrelation and global collinearity on GWR and MGWR in terms of potential local collinearity issues are similar to what have been observed for the smaller sample scenario (Tables 4 and 5). It is negligible for MGWR even in the extreme scenario where both correlation is very strong ($\rho = 0.95$ and $r = 0.95$).

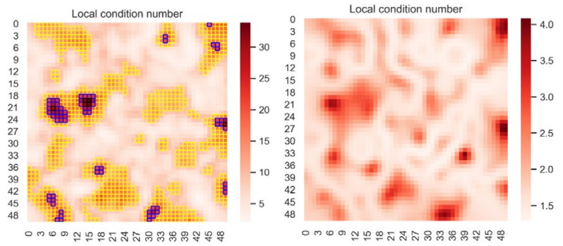
An examination of the scenario where both very strong spatial autocorrelation ($\rho = 0.95$) and collinearity (r=0.9) exist shows that this leads to a situation where the two covariate surfaces have similar spatial patterns (Fig. 16a). Figure 16b visualizes the spatial patterns of LCNs for both GWR and MGWR with locations that have an LCN larger than 30 highlighted with black boundaries, those with LCNs between 20 and 30 highlighted with blue boundaries, and those with LCNs between 10 and 20 highlighted with yellow boundaries. GWR produces many locations with high LCNs, especially where hot or cold spots coincide in the surfaces, due to the single bandwidth that is used for local estimation. This results in predominantly statistically insignificant parameter estimates and those that are not largely inflated compared to the true values. In contrast, MGWR does not produce any large LCNs (Fig. 16b), still recovers spatial patterns of true parameter surfaces, and produces

Table 5 Average percentage of locations that have a local condition number > 10 using MGWR (n = 2500)

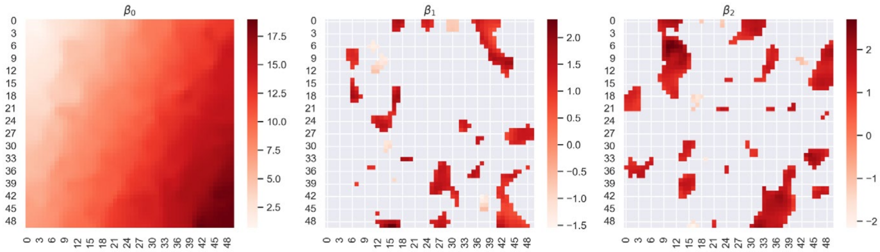
		Collinearity (r)					
		0.00	0.50	0.80	0.85	0.90	0.95
Spatial autoregressive parameter (ρ)	0.00	0.0	0.0	0.0	0.0	0.0	0.0
	0.50	0.0	0.0	0.0	0.0	0.0	0.0
	0.80	0.0	0.0	0.0	0.0	0.0	0.02
	0.85	0.0	0.0	0.0	0.0	0.0	0.16
	0.90	0.0	0.0	0.0	0.0	0.0	0.44
	0.95	0.01	0.0	0.01	0.03	0.22	2.4



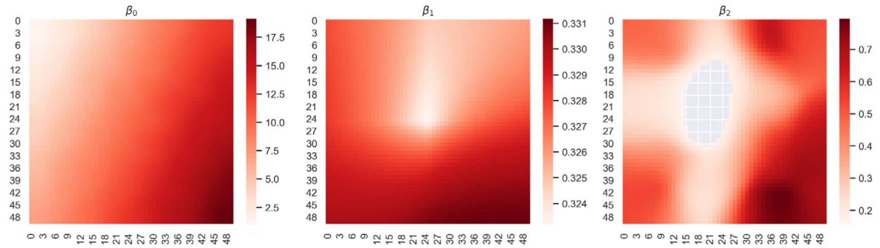
(a) Simulated Covariates



(b1) LCNs for GWR (b2) LCNs for MGWR

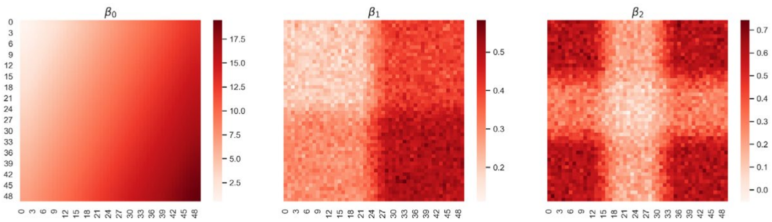


(c) GWR parameter estimates



(d) MGWR parameter estimates

Fig. 16 Results for MGWR and GWR for realization 18 under the scenario $\rho=0.95$ and $r=0.9$ ($n=2500$)



(e) "True" parameter surfaces

Fig. 16 (continued)

three unique optimal bandwidths (50, 2499, and 294 with 95% confidence intervals [50, 50], [2141, 2499], and [264, 316], respectively). However, similar to Fig. 15, the surface for β_1 in Fig. 16d has little-to-no overall variation compared to the true surface (i.e., an average effect) and is statistically insignificant according to a Monte Carlo test (p-value = 1.0). This further highlights the trade-off between overestimating the bandwidth and avoiding potential local collinearity issues. The outcome is that the true local effect is not identified, but in exchange, the overall model remains robust.

5 Discussions and conclusions

Two experimental designs were adopted to systematically explore the relationship between process scale (as measured by estimated optimal bandwidth values) and the levels of two types of correlation among explanatory variables (i.e., covariates) in (M)GWR, namely bivariate correlation (i.e., collinearity at a global level) and spatial autocorrelation. Previous work investigated the role of either collinearity or spatial autocorrelation but not their joint impact, and this was only carried out using the single-scale framework of GWR. By investigating these two characteristics of model input within the multiscale MGWR framework, it becomes possible to understand their effect on the estimated bandwidths and how this may potentially relate to the measured scale for each independent factor. The experiments covered a range of scenarios, including different sample sizes, noise levels, patterns of local processes (i.e., parameter surfaces), and levels of correlations. As a result, this work contributes additional evidence toward previous findings and generates several novel insights.

In concordance with previous findings, collinearity alone does not generally present any major issues until very high values. Taking reasonable precautions to only include variables that are not highly collinear should remain standard practice, and it is also important to use local extensions of diagnostics, such as LCNs, to identify any collinearity that could occur in local samples. Similarly, spatial autocorrelation alone does not seem to be an issue until very high levels. However, collinearity and spatial autocorrelation appear to have a joint effect that is stronger than either alone toward (a) generating local collinearity (as measured by the LCN) and (b) causing estimation error. In addition, removing one or more covariates because they are highly collinear may result in fewer explanatory factors, but because there is redundant information, the overall loss of explained variation should not be heavily affected. In contrast, highly spatially autocorrelated covariates are often strong predictors within a model and removing them eliminates a unique contribution to the model (unless there is also strong collinearity present). As a result, focusing more on global *and* local collinearity is still likely a sufficient management strategy, though it is useful to be aware that the general effect of collinearity can more quickly become accentuated if there are many variables with high spatial autocorrelation.

Furthermore, novel evidence was produced to support that MGWR provides significant improvements over GWR in terms of avoiding potential issues that might arise from collinearity and/or spatial autocorrelation. In fact, compared to GWR, MGWR typically produces less estimation error (overall and pattern recovery) and little-to-no potential local collinearity issues due to either collinearity, autocorrelation, or both until extreme levels, which would be easily defended against using typical diagnostics. These results are encouraging, providing additional evidence of the robustness of MGWR to potential collinearity issues (Oshan et al. 2020), highlighting the benefits of allowing the measurement of scale to vary for each relationship, and generally supporting the use of MGWR over GWR.

Moving to the multiscale paradigm provided the additional opportunity to investigate the role of collinearity and autocorrelation in measuring process scale. By examining the optimally estimated bandwidth values across each simulation design, three important trends emerged. First, the positive correlation between the noise level and the optimal MGWR bandwidths, along with the preserved order of bandwidth magnitudes, suggests that MGWR bandwidths should likely be interpreted in a relative rather than absolute manner. Second, in the design with a smaller sample, where the true parameter surface was global (i.e., stationary) with no spatial variation, the simultaneous presence of high spatial autocorrelation and collinearity in covariates caused the bandwidth for parameters with the lowest level of spatial heterogeneity (including the intercept) to be underestimated. This led to the detection of locally varying effects even when none existed, resulting in a form of false positive. However, the patterning of these local effects was not very pronounced. This finding serves as a cautionary note, particularly when interpreting the estimated local intercept as the measurement of intrinsic contextual effects (Fotheringham et al. 2021; Fotheringham and Li 2023). In such scenarios, identifying and managing covariate spatial autocorrelation and multicollinearity becomes even more critical. A third trend was that when collinearity becomes extreme (if it is allowed to do so), MGWR's ability to use different bandwidths enables it to adapt by selecting a larger bandwidth to increase the sample size, thereby mitigating the negative effects of smoothing. This implies a trade-off between missing the local effect for one covariate and keeping the model robust for other covariates—essentially a false negative, but the outcome is a near-stationary (i.e., global) estimate that approximates an average of the true local parameters. The result is that MGWR enhances overall model, while also suggesting that very large bandwidths may indicate either a global process or the presence of multicollinearity if caution is not exercised. Future work should further explore the extent to which pockets of local collinearity may trigger this effect and its implications for multiscale modeling.

The experiments here were simplified in the sense that there were only two explanatory variables. A key finding from the supplementary experiments, which introduced an additional independent covariate, is that spatial autocorrelation and collinearity primarily affect only the associated parameter and bandwidth estimation (and the intercept in the case of spatial autocorrelation), assuming all relevant variables are included. This confirms the isolating effect of these correlations on specific model components. However, in reality, covariates are more likely than not to exhibit both spatial autocorrelation and associations with each other. As the number of covariates increases, the effects of collinearity and spatial autocorrelation are expected to compound more quickly, particularly at the local level, where detection may require additional effort. Although using MGWR may help mitigate some of these issues, caution is warranted regarding false positives on the local intercept and false negatives on local slopes. The latter is not inherently problematic for the rest of the model and is just an extension of the typical problem with multicollinearity—a variable that is collinear might be removed to avoid obfuscating other relationships in a regular regression, whereas in a local regression, a variable might not need to be removed but cannot have a local effect without obfuscating other relationships. As a result, a local effect might be missed, but the remaining relationships may remain robust.

Including this current work, the notion of multicollinearity in GWR has been studied now for approximately two decades (Wheeler and Tiefelsdorf 2005; Wheeler 2006; Fotheringham and Oshan 2016). As more data availability expands and methodologies continue to evolve, revisiting the matter offers an opportunity to develop a more nuanced understanding. Specifically, with the advancement of GWR to MGWR (Fotheringham et al. 2017; Yu et al. 2020a), the assumption of a constant bandwidth is relaxed to provide insights into the spatial scale at which each process operates. Larger bandwidths indicate global processes and less smoothing of the data to create local samples, while smaller bandwidths indicate local processes and more smoothing of the data. It is within this context that this work was motivated to reprise the topic and consequently contribute to a more comprehensive understanding of MGWR, highlighting considerations for effectively using the method for empirical applications.

As with any simulation studies, this research has limitations. First, the simulations only consider two collinear covariates, limiting the ability to assess more complex correlations involving multiple covariates. Second, when collinearity is present, both covariates are assumed to exhibit the same level of spatial autocorrelation. Allowing them to have different spatial autocorrelation levels could provide additional and meaningful insights. Third, only regular spatial configuration is examined, while empirical studies are often faced with irregular spatial configurations. Fourth, local spatial autocorrelation could exist even in the absence of significant global spatial autocorrelation in the explanatory variables, an aspect that remains unaddressed. Future research will aim to overcome these limitations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10109-025-00468-1.

Declarations

Conflict of interest This research was supported in part by National Science Foundation Grant TI-2345820 and CMMI-2151970.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht
Anselin L, Amaral P (2023) Endogenous spatial regimes. J Geogr Syst. https://doi.org/10.1007/s10109-023-00411-2
Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data, 2nd edn. CRC Press, Boca Raton
Fotheringham AS (2023) A comment on 'a route map for successful applications of geographically-weighted regression': the alternative expressway to defensible regression-based local modeling. Geogr Anal 55(1):191–197
Fotheringham AS, Li Z (2023) Measuring the unmeasurable: models of geographical context. Ann Am Assoc Geogr 113(10):2269–2286
Fotheringham AS, Oshan TM (2016) Geographically weighted regression and multicollinearity: dispelling the myth. J Geogr Syst 18(4):303–329
Fotheringham AS, Charlton ME, Brunsdon C (1998) Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environ Plan A 30(11):1905–1927
Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, Hoboken
Fotheringham AS, Brunsdon C, Charlton M (2003) Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, Hoboken
Fotheringham AS, Yang W, Kang W (2017) Multiscale geographically weighted regression (MGWR). Annals Am Assoc Geogr 107(6):1247–1265. https://doi.org/10.1080/24694452.2017.1352480
Fotheringham AS, Hanchen Yu, Wolf LJ, Oshan TM, Li Z (2022) On the notion of 'bandwidth' in geographically weighted regression models of spatially varying processes. Int J Geogr Inf Sci IJGIS 36(8):1485–1502
Fotheringham AS, Oshan TMM, Li Z (2023) multiscale geographically weighted regression: theory and practice. CRC Press, Boca Raton
Gelfand AE, Hyon-Jung Kim CF, Sirmans SB (2003) Spatial modeling with spatially varying coefficient processes. J Am Stat Assoc 98(462):387–396. https://doi.org/10.1198/016214503000170
Geniaux G, Martinetti D (2018) A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. Reg Sci Urban Econ 72:74–85
Goldberger AS (1964) Econometric theory. Wiley, New York
Goldberger AS (1991) A course in econometrics. Harvard University Press, Cambridge MA
Griffith D, Chun Y, Li B (2019) Spatial regression analysis using eigenvector spatial filtering. Academic Press, Cambridge
Hu Y, Binbin Lu, Ge Y, Dong G (2022) Uncovering spatial heterogeneity in real estate prices via combined hierarchical linear model and geographically weighted regression. Environ Plann b Urban Anal City Sci 49(6):1715–1740

Li K, Lam NSN (2018) Geographically weighted elastic net: a variable-selection and modeling method under the spatially nonstationary condition. Annals Am Assoc Geogr 108(6):1582–1600. https://doi.org/10.1080/24694452.2018.1425129

Li AZ, Fotheringham S, Oshan TM, Wolf LJ (2020) Measuring bandwidth uncertainty in multiscale geographically weighted regression using akaike weights. Annals Am Assoc Geogr 110(5):1500–1520. https://doi.org/10.1080/24694452.2019.1704680

Murakami D, Yoshida T, Seya H, Griffith DA, Yamagata Y (2017) A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. Spat Stat 19:68–89

Oshan TM (2023) Navigating the methodological landscape in spatial analysis: a comment on "a route map for successful applications of geographically-weighted regression." Geogr Anal 55(1):179–183

Oshan TM, Li Z, Kang W, Wolf LJ, Stewart Fotheringham A (2019) Mgwr: a python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. ISPRS Int J Geo Inf 8(6):269

Oshan TM, Smith JP, Stewart Fotheringham A (2020) Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. Int J Health Geogr 19(1):11

Páez A, Farber S, Wheeler D (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environ Plan A 43(12):2992–3010

Stewart Fotheringham A, Li Z, Wolf LJ (2021) Scale, context, and heterogeneity: a spatial analytical perspective on the 2016 US presidential election. Annals Am Assoc Geogr 111(6):1–20

Sugasawa S, Murakami D (2021) Spatially clustered regression. Spat Stat 44:100525

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234. https://doi.org/10.2307/143141

Wheeler DC (2009) Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted Lasso. Environ Plan A 41(3):722–742

Wheeler D, Tiefelsdorf M (2005) Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J Geogr Syst 7(2):161–187

Wheeler DC (2006) Diagnostic tools and remedial methods for collinearity in linear regression models with spatially varying coefficients

Wheeler DC (2010) Visualizing and diagnosing coefficients from geographically weighted regression models. In: Geospatial analysis and modelling of urban structure and dynamics, pp 415–36.

Wolf LJ, Oshan TM, Stewart Fotheringham A (2018) Single and multiscale models of process spatial heterogeneity. Geogr Anal 50(3):223–246

Yu H, Stewart Fotheringham A, Li Z, Oshan T, Kang W, Wolf LJ (2020a) Inference in multiscale geographically weighted regression. Geogr Anal 52(1):87–106

Yu H, Stewart Fotheringham A, Li Z, Oshan T, Wolf LJ (2020b) On the measurement of bias in geographically weighted regression models. Spat Stat 38:100453