

RESEARCH ARTICLE

Hyper-local geographically weighted regression: extending GWR through local model selection and local bandwidth optimization

Alexis Comber*¹, Yunqiang Wang*², Yihe Lü³, Xingchang Zhang⁴, and Paul Harris⁵

¹School of Geography, University of Leeds, UK

²State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China

³State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences; Joint Center for Global Change Studies; University of Chinese Academy of Sciences, Beijing, China

⁴Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, China

⁵Sustainable Agricultural Sciences—North Wyke, Rothamsted Research, UK

Received: February 7, 2018; returned: May 31, 2018; revised: June 4, 2018; accepted: July 9, 2018.

* co-corresponding authors

Abstract: Geographically weighted regression (GWR) is an inherently exploratory technique for examining process non-stationarity in data relationships. This paper develops and applies a *hyper-local* GWR which extends such investigations further. The hyper-local GWR simultaneously optimizes both local model selection (which covariates to include in each local regression) and local kernel bandwidth specification (how much data should be included locally). These are evaluated using a measure of model fit. The hyper-local GWR approach evaluates different kernel bandwidths at each location and selects the most parsimonious local regression model. By allowing models and bandwidths to vary locally, this approach extends and refines the one-size-fits-all “whole map model” and “constant bandwidth calibration” under standard GWR. The results provide an alternative, complementary and more nuanced interpretation of localized regression. The method is illustrated

using a case study modeling soil total nitrogen (STN) and soil total phosphorus (STP) from data collected at 689 locations in a watershed in Northern China. The analysis compares linear regression, standard GWR, and hyper-local GWR models of STN and STP and highlights the different locations at which covariates are identified as significant predictors of STN and STP by the different GWR approaches and the spatial variation in bandwidths. The hyper-local GWR results indicate that the STN processes are more non-stationary and localized than found via a standard application of GWR. By contrast, the results for STP are more confirmatory (i.e., similar) between the two GWR approaches providing extra assurance about the nature of the moderate non-stationary relationships observed. That is, a standard GWR may underestimate localized spatial heterogeneity where it is strongly present (as in the STN case study) and may overestimate it where spatial homogeneity is present (as in the STP case study). The overall benefits of hyper-local GWR are discussed, particularly in the context of the original investigative aims of GWR. A hyper-local approach provides a useful counter view of local regression modeling to that found with standard GWR. Where spatial non-stationarity exists, the hyper-local GWR provides a more spatially nuanced indication of the localization than a standard GWR analysis and can be used to suggest the direction of further analyses and investigations. Some areas of further work are suggested.

Keywords: Loess Plateau, geographically weighted regression, GWR, model selection, spatial analysis

1 Introduction

Geographically Weighted Regression (GWR), as first described in Brunson et al. [5], is a commonly used approach in spatial analysis. It has at its core the idea that global or whole map statistical models may make unreasonable assumptions of spatial non-stationarity amongst the processes under investigation [33]. The intention of GWR was to provide an exploratory approach to investigate the spatial nature of relationships between response and predictor variables, and in so doing, to provide a better understanding of the process under consideration. It has conceptual elegance; local regression models are constructed at different locations using data under a moving window or kernel, which are weighted by the distance to the kernel center such that data furthest away contribute less to the overall model. Because of this, the *geographically weighted* (GW) framework has been extended to include different types of models including GW principal components analysis [20], GW summary statistics [4], GW discriminant analysis [6], GW variograms [22], GW Structural Equation Models [11], and has been applied in domains with little tradition of local statistical approaches such as remote sensing (eg [9, 12, 15]). The fundamental aims of GWR and GW frameworks are thus to explore spatial relationships in data and processes.

One of the key parts of any GW analysis is to determine an optimal kernel size or bandwidth, as this controls how much data are included in each local model and the degree of smoothing or “localness” in the GW model. Gollini et al. [19] provide a full discussion but in essence the bandwidth determines the scale at which each localized model operates. Smaller bandwidths result in greater local variation in the outputs and larger ones result in outputs that are increasingly closer to the global measure. Optimum kernel bandwidths can be found by minimizing a model fit diagnostic and most GWR implementations use

a leave-one-out cross-validation (CV) score, the Akaike Information Criterion (AIC) [1], or a corrected version of the AIC [25]. Essentially what these do is construct a local model at each location for each bandwidth and then the model fit is calculated from all of the local models for that bandwidth. The bandwidth with the best (lowest) score is selected. A standard implementation of GWR frequently determines which covariates to include using a global model selection procedure and then determines the optimal bandwidth using the model fit procedure described above. A GWR generates coefficient estimates at each location and these are commonly mapped to show the spatial variation in the degree to which changes in covariates x are associated with changes in y . Thus bandwidth optimization and model selection are both global in nature, the same covariates and bandwidth are specified for each local regression of GWR.

This paper proposes an enhancement to standard applications of GWR that allows both model selection and bandwidth to vary locally. The aim of such *hyper-local* approaches to GWR is to provide a still deeper understanding of the spatial nature of the processes under investigation. As with GWR, hyper-local GWR applies a local regression under a moving window or kernel at each location under consideration, but it simultaneously optimizes both the local regression model and the local kernel bandwidth. This is entirely novel: although model selection in GWR has been done [42], it has not been combined with non-constant bandwidth selection where bandwidths are truly local and unique (i.e. [34, 35]). Local model selection helps to identify *which* covariates are important in explaining the variation in the dependent variable and *where* they are important. The corresponding local bandwidths in turn provide insight into the local scales of influence. The hyper-local GWR approach provides an alternative interpretation of localized regression by extending GWR through local model selection and local bandwidth optimization. It complements and enriches a standard application of GWR.

2 Methods

Linear regression, GWR, and the proposed hyper-local GWR were used to construct models of soil total nitrogen (STN) and soil total phosphorus (STP). The analyses used the data described in Wang et al. [41].

2.1 Data and study area

The case study data reports measurements made at 689 locations in the Liudaogou watershed, within the Loess Plateau, located 14 km West of Shenmu, Shaanxi Province, China. Wang et al. [41] provide a full description of the data including descriptive statistics and correlations amongst the variables used in the study. In brief, this is a small watershed with an altitudinal range of 1081m to 1274m, a semi-arid climate with mainly grassland land use. The data were collected at locations on an approximate 100m by 100m grid (Figure 1) and analyzed in the laboratory to provide measurements of covariates commonly associated with STN and STP: soil organic carbon (SOCgkg), clay (ClayPC), silt (SiltPC), sand (SandPC), nitrate nitrogen (NO₃Ngkg), and ammonium (NH₄Ngkg). Some of the variables were transformed using natural logs (STN, SOCgkg, NO₃Ngkg, NH₄Ngkg) and square roots (STP, ClayPC), as was done by [41].



Figure 1: The sample locations and some context from the OpenStreetMap Bing layer.

2.2 Linear regression and GWR

A standard linear regression for spatial data is specified as follows:

$$y_i = \beta_0 + \sum_{j=1}^m \beta_j x_{ij} + \epsilon_i \quad (1)$$

where for observations indexed by $i = 1, \dots, n$, y_i is the response variable, x_{ij} is the value of the j^{th} predictor variable, m is the number of predictor variables, β_0 is the intercept term, β_j is the regression coefficient for the j^{th} predictor variable and ϵ_i is the random error term. GWR is similar in form to linear regression, except that GWR calculates a series of local linear regressions rather than one global one. A GWR model has locations associated with the coefficient terms:

$$y_i = \beta_{0(u_i, v_i)} + \sum_{j=1}^m \beta_{j(u_i, v_i)} x_{ij} + \epsilon_i \quad (2)$$

where (u_i, v_i) is the spatial location of the i^{th} observation and $\beta_{j(u_i, v_i)}$ is a realization of the continuous function $\beta_j(u, v)$ at point i . The geographical weighting results in data nearer to the kernel center making a greater contribution to the estimation of regression coefficients at each local regression calibration point k . For this study, the weights were generated using a *bisquare* kernel for the bandwidth parameter which is defined by:

$$w_{ik} = (1 - (d_{ik}/r_k)^2)^2 \text{ if } d_{ik} \leq h, w_{ik} = 0 \text{ otherwise} \quad (3)$$

where d_{ik} is the distance between the kernel centre and regression calibration point k and h is the bandwidth. Here h can be specified as a fixed (constant) distance value, or in an

adaptive, varying distance way, where the number of nearest neighbors is fixed (constant). In this case, fixed, distance-based kernel bandwidths were determined using the AIC-based model fit procedure. Fixed bandwidths were chosen to support direct understandings of the spatial scales of relationship non-stationarity and because the data locations are regularly spaced.

2.3 Hyper-local GWR

In a hyper-local GWR, both the bandwidth and the regression model selection are optimized locally rather than globally across all local models as in a standard GWR. A sequence of bandwidths was investigated (from 200 m to 3700 m in steps of 50 m, $n = 63$) and at each location regression models of STN and STP were constructed using weighted data falling under the kernel. Then a stepwise AIC model selection procedure was applied—in this case the *stepAIC* function in the MASS R package [36]. Thus for each location, 63 local regression models of STN and STP were constructed, and a stepwise AIC was used to determine the locally selected regression model at each location under each kernel size. The AIC scores of each selected model was calculated resulting in 63 AIC scores at each location. The “best” model and bandwidth combination at each location was that with the lowest AIC. The use of AIC as a method for model selection will be returned to in the discussion.

3 Results

3.1 Linear regression

Linear regression models of STN and for STP were constructed from the six covariates and a stepwise AIC model selection procedure was applied. Tables 1 and 2 summarize the coefficient estimates and the selected covariates.

	Full			Selected		
	Estimate	Std. Error	Pr(> t)	Estimate	Std. Error	Pr(> t)
(Intercept)	-3.823	1.134	0.001	0.011	0.314	0.754
SOCgkg	0.688	0.040	0.000	0.050	6.762	0.000
ClayPC	0.081	0.084	0.337	.	.	.
SiltPC	0.028	0.010	0.004	0.004	4.678	0.000
SandPC	0.016	0.011	0.150	.	.	.
NO3Ngkg	0.125	0.029	0.000	0.002	1.936	0.053
NH4Ngkg	-0.138	0.074	0.061	.	.	.
	$R^2:0.610$	adj $R^2:0.607$,	AIC:1123.7	$R^2:0.141$	adj $R^2:0.137$	AIC:586.1

Table 1: Summary of the coefficient estimates arising from the Full and AIC selected linear regression models of STN.

In the case of STN, the full model is being driven by SOCgkg, SiltPC, and NO3Ngkg which are significantly associated with STN. The model is similar to that described in Wang et al. [41] with an R^2 of 0.61 and all of the covariates positively associated with STN except NH4Ngkg. The AIC selected model does not include the ClayPC, SandPC, and NH4Ngkg covariates. Observe that NO3Ngkg is not significantly associated with STN in the AIC selected model (at the 95% level). The significant predictors of STP in the full model were

	Full			Selected		
	Estimate	Std. Error	Pr(> t)	Estimate	Std. Error	Pr(> t)
(Intercept)	-0.078	0.163	0.632	0.201	0.014	0.000
SOCgkg	0.047	0.006	0.000	0.014	0.002	0.000
ClayPC	0.011	0.012	0.366	-0.005	0.002	0.002
SiltPC	0.007	0.001	0.000	0.004	0.000	0.000
SandPC	0.005	0.002	0.002	.	.	.
NO3Ngkg	-0.001	0.004	0.866	.	.	.
NH4Ngkg	0.026	0.011	0.013	0.001	0.001	0.023
	R^2 :0.404	adj R^2 :0.399	AIC:-1548.2	R^2 :0.326	adj R^2 :0.322	AIC:-1178.9

Table 2: Summary of the coefficient estimates arising from the Full and AIC selected linear regression models of STP.

SOCgkg, SiltPC, SandPC, and NH4Ngkg with an R^2 value of 0.40, again similar to the findings of Wang et al (2009). The selected model did not include the covariates for NO3Ngkg and SandPC, but all retained covariates were significant. In both cases the selected models reflect the impact of silt and soil organic carbon in increasing the soil surface area supporting higher absorption capacities, and thus concentrations of STN and STP, as noted by Wang et al. [41]. The AIC selected models are more parsimonious model but with weaker R^2 and adjusted R^2 values as would be expected.

Note that the selected model does not necessarily include covariates that are significant and that non-significant covariates in the full model may be included in the selected model and may become significant (e.g., the ClayPC covariate for the STP regression). The key point is that the variance in STN and STP can be explained by two competing, but equally valid linear regression models. This concept is repeated locally in the subsequent GWR analyses and is a cornerstone of this paper.

3.2 Standard GWR

Linear regression models assume that the contributions to the model made by the different covariates are the same across the study area. In reality, this assumption of process spatial invariance may be violated and GWR seeks to quantify the spatial variation in data relationships. In a standard GWR analysis, covariate selection is typically undertaken globally and the same regression model is constructed locally using weighted data subsets. The coefficient estimates are commonly mapped and local covariate selection (and goodness of fit evaluations) can be done by identifying local covariate t -values that indicate coefficients to be significantly different from zero (e.g., [24]).

The optimal bandwidths for GWR models of STN and STP were found at 1026m and 1629m, respectively. These were used to calibrate the GWRs constructed at each of the sample locations in Figure 1. The local coefficient estimates from these are summarized in Tables 3 and 4. The GWR coefficients for STN show considerable spatial variation (via the inter-quartile range, IQR) and much less is found in the local STP models, as also reflected in the larger bandwidth. For example, in the STN GWR model the coefficient estimates for SandPC and NO3Ngkg have IQRs of 0.0408 and 0.1267, respectively, while in the STP GWR model these have relatively small IQRs (0.0023 and 0.0077, respectively). However

note the relatively high variation of the IQRs of the local coefficient estimates in the GWR models compared to the global coefficient estimates.

	1st Qu.	Median	Mean	3rd Qu.	IQR	Global
Intercept	-4.9409	-2.8261	-3.0396	-0.8880	4.0529	-3.8229
SOCgkg	0.6137	0.6730	0.6760	0.7509	0.1372	0.6882
ClayPC	-0.0162	0.0865	0.0752	0.1630	0.1792	0.0811
SiltPC	0.0012	0.0166	0.0201	0.0390	0.0378	0.0284
SandPC	-0.0092	0.0068	0.0099	0.0315	0.0407	0.0156
NO3Ngkg	0.0399	0.0909	0.1282	0.1666	0.1267	0.1247
NH4Ngkg	-0.2313	-0.1115	-0.1661	-0.0395	0.1918	-0.1384

Table 3: The distributions of the coefficient estimates arising from a GWR model of STN.

	1st Qu.	Median	Mean	3rd Qu.	IQR	Global
Intercept	-0.2119	-0.1046	-0.0683	0.0640	0.2759	-0.0781
SOCgkg	0.0371	0.0410	0.0422	0.0487	0.0116	0.0469
ClayPC	-0.0130	0.0132	0.0063	0.0261	0.0391	0.0110
SiltPC	0.0059	0.0081	0.0078	0.0098	0.0039	0.0074
SandPC	0.0038	0.0050	0.0051	0.0061	0.0023	0.0049
NO3Ngkg	-0.0016	0.0027	0.0031	0.0061	0.0077	-0.0007
NH4Ngkg	0.0000	0.0186	0.0180	0.0412	0.0412	0.0263

Table 4: The distributions of the coefficient estimates arising from a GWR model of STP.

The spatial variations in the coefficient estimates arising from the two GWR models are mapped in Figures 2 and 3 and indicate the relative importance of the contribution made to each local model by each covariate at each location. They confirm that there is much greater spatial variation in the relationships associated with STN than with STP.

The *t*-values in Figures 2 and 3 show where local coefficients are significant and thus where a covariate is an important predictor of STN or STP. This provides an indication of local covariate selection from the full model and is analogous to the global full models reported in Tables 1 and 2. For example, it is evident in both GWR models that SOCgkg is strongly and significantly associated with STN and STP across all locations, but the strength of this association varies spatially. Whereas significant coefficient estimates of NO3Ngkg are highly localized in each GWR model indicating strong associations in the north east and center of the study area with STN and strong associations in the north with STP. In general, significant relationships are much more localized for STN than for STP.

3.3 Hyper-local GWR

The GWR analysis applied the same kernel bandwidth and included the same full set of covariates in each local regression model. Figures 2 and 3 display the spatial distribution of the GWR coefficient estimates and a degree of local model selection is possible through exploration of the local *t*-values associated with the local coefficient estimates. This is a standard application of GWR, supporting investigations of process heterogeneity with respect to spatially-varying relationships.

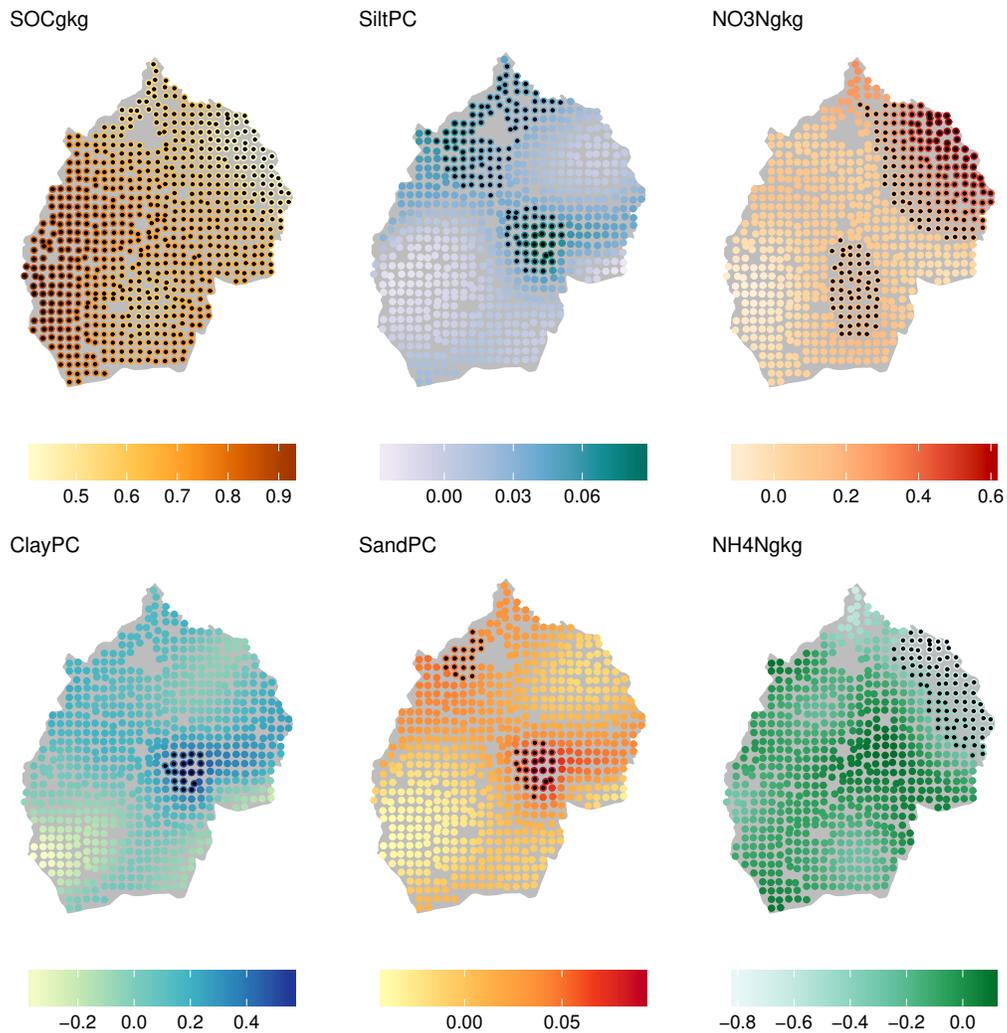


Figure 2: Spatial variation in coefficient estimates from a standard GWR model of STN. Significant t -values are indicated by the black shaded points.

The hyper-local GWR approach provides an alternative interpretation of localized regression through local model selection and local bandwidth optimization. It builds on previous GWR studies by Paez and Wheeler that have identified analytical advantages when locally-determined, non-constant bandwidths are applied [34,35] and when covariate selection is determined locally [42]. It combines these localized characteristics but the ultimate objective is entirely different to the studies by Paez and Wheeler. Paez et al. [34, 35] were concerned about modeling a non-stationary error variance in GWR via a parametric approach and Wheeler [42] sought to address local collinearity in GWR via a lasso approach.

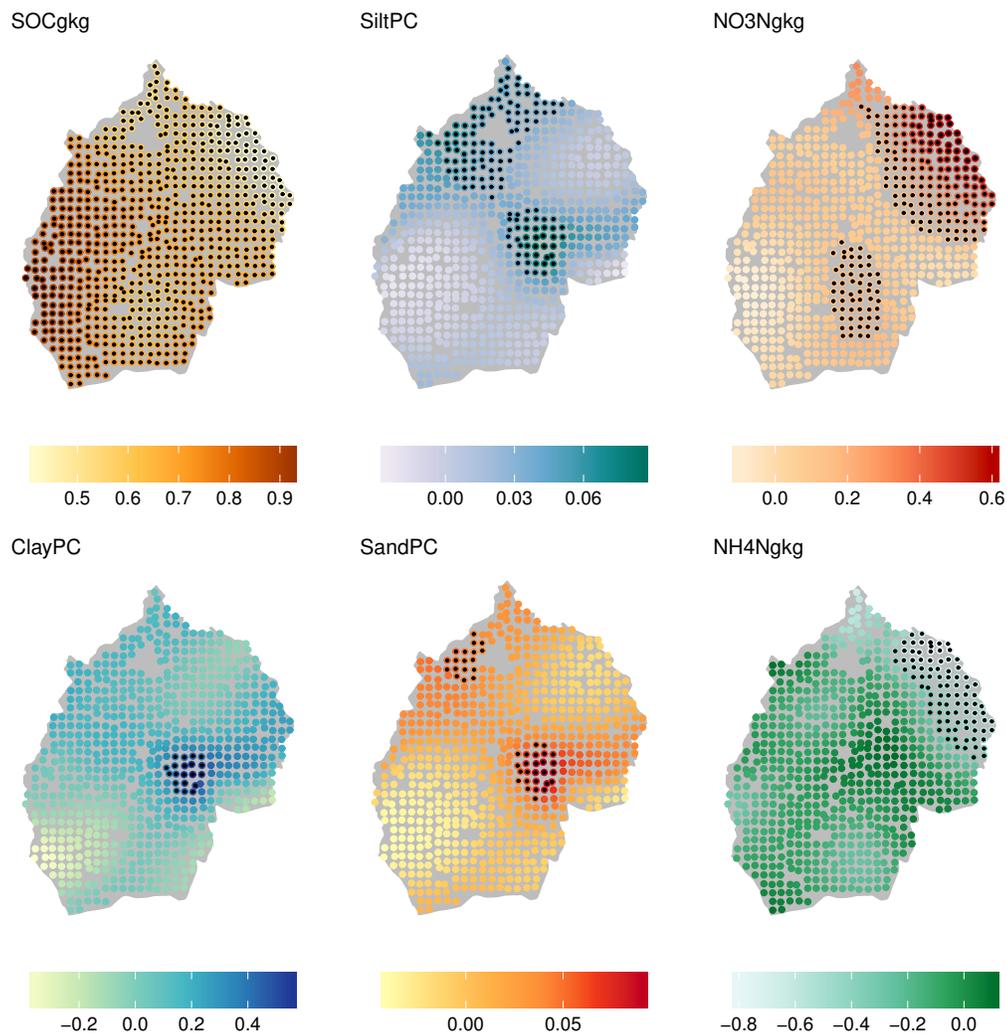


Figure 3: Spatial variation in coefficient estimates from a standard GWR model of STP. Significant t -values are indicated by the black shaded points.

For each of the 689 data points, the hyper-local GWR identified the components of the best fitting model for each of the 63 bandwidths (from 200 m to 3700 m in intervals of 50 m) and returned the AIC score for the model. Thus it was possible to determine the best fitting model, with the lowest AIC score at each location.

3.3.1 Local bandwidth selection

Figure 4 shows bandwidths with the lowest AIC scores from the hyper-local GWR models of STN and STP. They exhibit different spatial patterns and characteristics. The STN bandwidths range from 200-1800 m, with larger bandwidths (say, 1000-1800 m) traversing

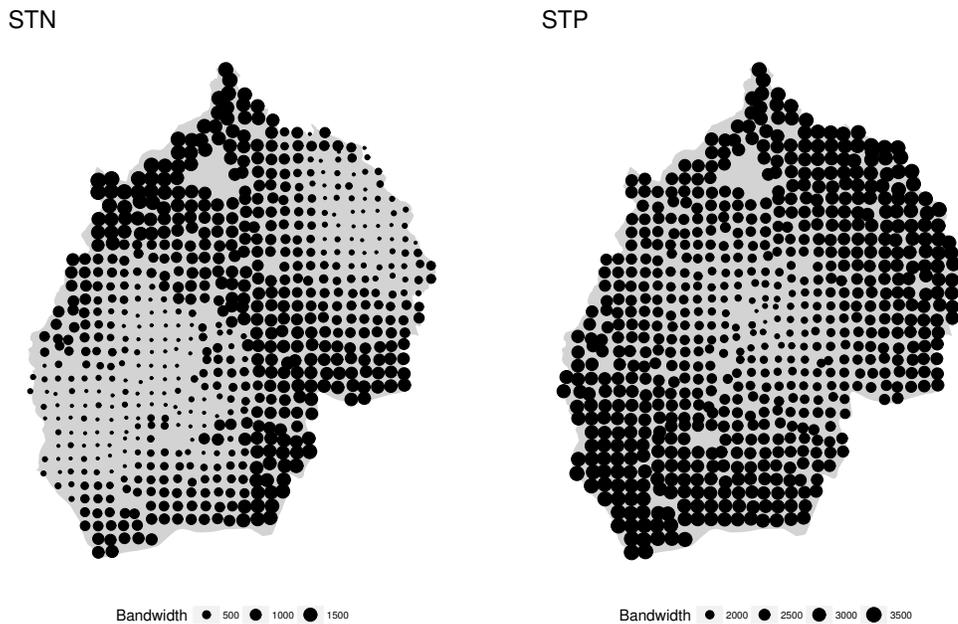


Figure 4: Spatial variation in local bandwidth size (in metres) of the hyper-local GWR models of STN and STP.

from the southeast to the northwest. This suggests that local regressions in this area are informed by data subsets of a similar size to that found with standard GWR (with its constant bandwidth of 1026 m). Elsewhere, the bandwidths are much smaller (200-1000 m), so that local regressions in these areas are informed by much smaller data subsets. The distribution of bandwidths in the hyper-local GWR model is on the whole indicative of increased localized spatial heterogeneity in data relationships, which is more than that suggested by the standard GWR analyses above.

Conversely, the STP bandwidths range from 1500-3700 m and are much larger almost everywhere than the constant bandwidth for standard GWR at 1629 m. Thus, most of the local regressions in a hyper-local GWR are informed by much larger data subsets than a standard GWR. Only to the center of the study area are bandwidths from hyper-local GWR of similar size to a standard GWR. The larger bandwidths indicate reduced spatial heterogeneity to that found with standard GWR, and suggests spatial homogeneity in the relationships (i.e., tending to the global regression).

3.4 Local covariate selection and distribution of coefficient t-values

Investigating the spatial variation in bandwidth size is only one aspect of hyper-local GWR and should be linked to consideration of local covariate selection. Table 5 summarizes how many times each covariate was selected using stepwise AIC at each of the 689 locations in the hyper-local GWR models. There are a number of interesting points. STN model selection for the global regression (Table 1 excluded ClayPC, SandPC, and NH4Ngkg, while



these are now selected in 522, 641, and 439 out of 689 hyper-local models, respectively). STP model selection for the global regression (Table 2) excluded SandPC and NO3Ngkg, while these are now selected in 689 and 170 out of 689 local models, respectively). Additionally, three covariates were always selected regression (SOCgkg, SiltPC, and SandPC), whereas for STN none were. This suggests that there are potentially interesting local interactions between covariates which are missed in standard GWR in which all six covariates are included in the model for all 689 local regressions.

	STN	STP
SOCgkg	657	689
ClayPC	522	527
SiltPC	505	689
SandPC	641	689
NO3Ngkg	476	170
NH4Ngkg	439	425

Table 5: Number of sample locations where different covariates were selected in hyper-local GWR.

Figures 5 and 6 show the spatial variation in the selection of covariates and their coefficient estimates in the hyper-local GWR models and where the coefficient estimates were found to be locally significant via their t -values. The coefficient estimates arising from the hyper-local GWR models are summarized in Tables 6 and 7. Note that the legends in Figures 5 and 6 display data values within the inter-quartile ranges of the hyper-local GWR coefficient estimates. Extreme values were set to the 25th or 75th quartile value, to provide a meaningful shading. Of interest is where the hyper-local GWR coefficients were found to be significant via their t -values and how these compare to the maps of t -values in Figures 2 and 3 for standard GWR. Some large local differences are evident, especially for STN.

For example, in the STN models (comparing Figures 2 and 5), SandPC is a significant covariate at most locations in the hyper-local GWR model. In the standard GWR model (Figure 2 it is only significant in two sub-regions to the north and center of the study area. Whilst, NH4Ngkg in the standard GWR model of STN is significant in the north-west of the study area, but has a much wider significance in the hyper-local GWR model. These results indicate that when the bandwidth and covariate selection are more localized under the hyper-local GWR, then significant non-stationary relationships result, that are not apparent with standard GWR. Similar interpretations apply to STN relationships with SiltPC, NO3Ngkg, and ClayPC while it appears that STN's relationship to SOCgkg is consistent across both GWR forms. Note also that hyper-local GWR tends to provide spatially disjoint areas of covariate selection and coefficient significance, reflecting highly localized processes. For the STP process, comparing Figures 3 and 6, there are very similar patterns for significant coefficients from the hyper-local GWR and from the standard GWR for all six covariates, although NO3Ngkg, SandPC, ClayPC, and NH4Ngkg show enlarged localized areas of significance under the hyper-local model. Note that NO3Ngkg is only selected in 170 sample locations in hyper-local GWR (see Table 5) and these are in the north, precisely where the standard GWR shows the NO3Ngkg relationships as significant.

Clearly, these results indicate that when the bandwidth and covariate selection tend towards the global solution, as with the hyper-local GWR of STP, the non-stationary relationships that result from a hyper-local GWR are broadly similar for both forms of GWR.

However, where localized spatial heterogeneity is present in data relationships, as with STN, the hyper-local GWR provides a more spatial nuanced indication of the localization than a standard GWR analysis.

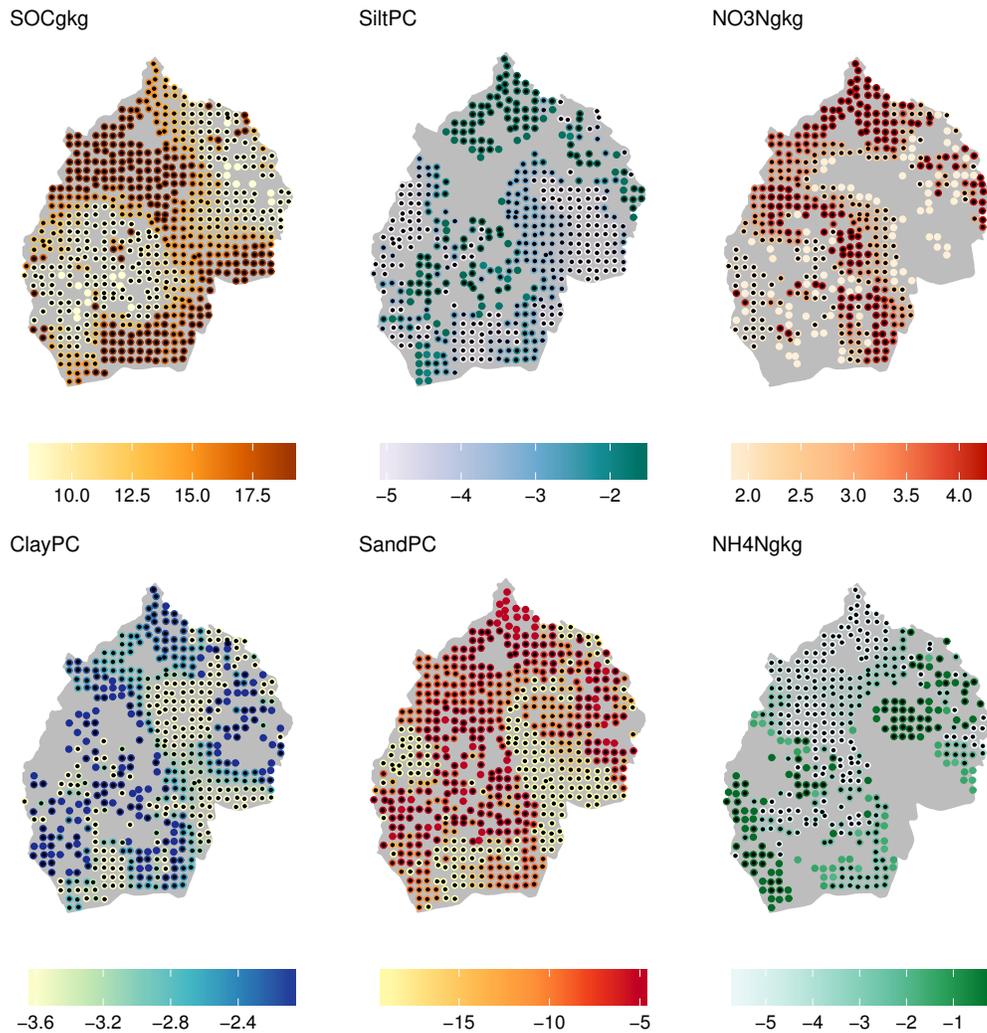


Figure 5: The spatial distribution of the selected covariates included in each hyper-local GWR model of STN. Significant t -values are indicated by the black shaded points.

3.5 Comparisons of global and local model fit

The final analysis compared the three different regression models in the degree to which they (in-sample) predict STN and STP. The scatterplots in Figure 7 show fitted values against observed values for these six models. For STN, the model fits improve with increas-

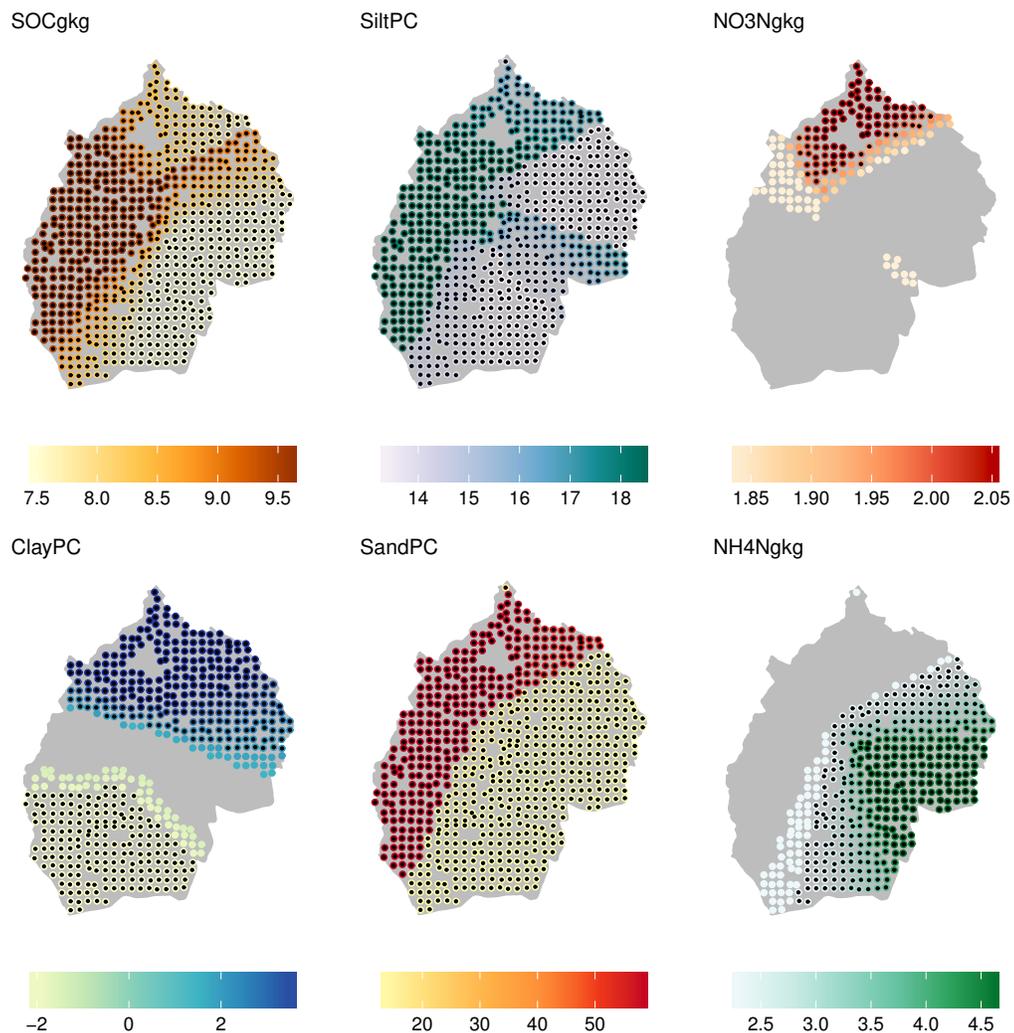


Figure 6: The spatial distribution of the selected covariates included in each hyper-local GWR model of STP. Significant t -values are indicated by the black shaded points.

ing spatial nuance, from linear regression (full model), to standard GWR and to hyper-local GWR (R^2 of 0.61, 0.68 and 0.94, respectively). Rather surprisingly, there is little improvement in model fit from the linear regression to standard GWR. The strong predictive performance of hyper-local GWR can be attributed to the local tightening of bandwidths and variable selection. For STP, the model fits do not improve in the same way. There is a moderate increase from linear regression (full model) to standard GWR but then a small decrease to the hyper-local GWR (R^2 of 0.40, 0.47, and 0.45 respectively). The decrease in R^2 observed for hyper-local GWR simply reflects that this model is actually not as local as

	1st Qu.	Median	Mean	3rd Qu.
SOCgkg	8.3827	13.7124	15.1081	19.1944
ClayPC	-3.6305	-2.7275	-3.8169	-2.0444
SiltPC	-5.1766	-3.2462	-4.0274	-1.4613
SandPC	-19.4445	-8.8947	-14.0557	-4.8087
NO3Ngkg	1.8363	3.1133	2.7784	4.3087
NH4Ngkg	-5.7028	-3.2104	-2.6720	-0.1863

Table 6: The distributions of the coefficient estimates arising from the hyper-local GWR model of STN.

	1st Qu.	Median	Mean	3rd Qu.
SOCgkg	7.4957	8.4218	8.5008	9.6309
ClayPC	-2.3965	1.8810	1.0432	3.9876
SiltPC	13.3632	15.3115	18.3168	18.4171
SandPC	12.1923	13.4488	31.7100	59.7112
NO3Ngkg	1.8337	1.9602	1.7673	2.0503
NH4Ngkg	2.2802	3.4096	3.5812	4.6855

Table 7: The distributions of the coefficient estimates arising from the hyper-local GWR model of STN.

standard GWR, as the bandwidths for hyper-local GWR tend to be larger and the process tends towards the global fit.

Care must be taken in the interpretation of model fit results, as any form of localized regression will tend to provide an improved prediction accuracy, the more complex it gets (hence the strong performance of hyper-local GWR for STN). Furthermore, although hyper-local GWR is shown to improve fit for the STN process, this has little predictive value, as hyper-local GWR cannot be used as an out-of-sample predictor. This is because the out-of-sample prediction does not have its own local bandwidth, whereas for standard GWR, the global bandwidth can be used [23]. Thus, hyper-local GWR is solely for guiding spatial exploration and inference only, as demonstrated in this study.

It is important to investigate local model fit characteristics so that the outputs in Figures 2 to 6 can be placed in better context and geographically contrasted. Figure 8 compares the local R^2 values for standard GWR and hyper-local GWR models for STN and STP and indicates that hyper-local GWR provides a better fit in 503/689 and 5/689 locations for STN and STP, respectively. Thus, for the STN process, the local regressions of standard GWR could be considered sub-optimal in 73% of the locations, whilst for the STP process, the local regressions of standard GWR are, in general, reasonable. The magnitude of the differences are much greater for STN than for STP. If Figure 8 is compared with Figure 4, the areas where a hyper-local approach provides a better model fit for STN directly correspond to those where a much smaller local bandwidth was selected. This behavior is not so apparent for the STP process.

The maps in Figure 8 confirm what has already been described. For STN, hyper-local GWR suggests a more localized relationship process where local model fit can improve using fewer data points and fewer covariates. Standard GWR is under-fitting the true non-stationary relationship process and this effect is not uncommon (e.g. [21]). Conversely, it is always possible that hyper-local GWR is overfitting. The STN process is, in general,

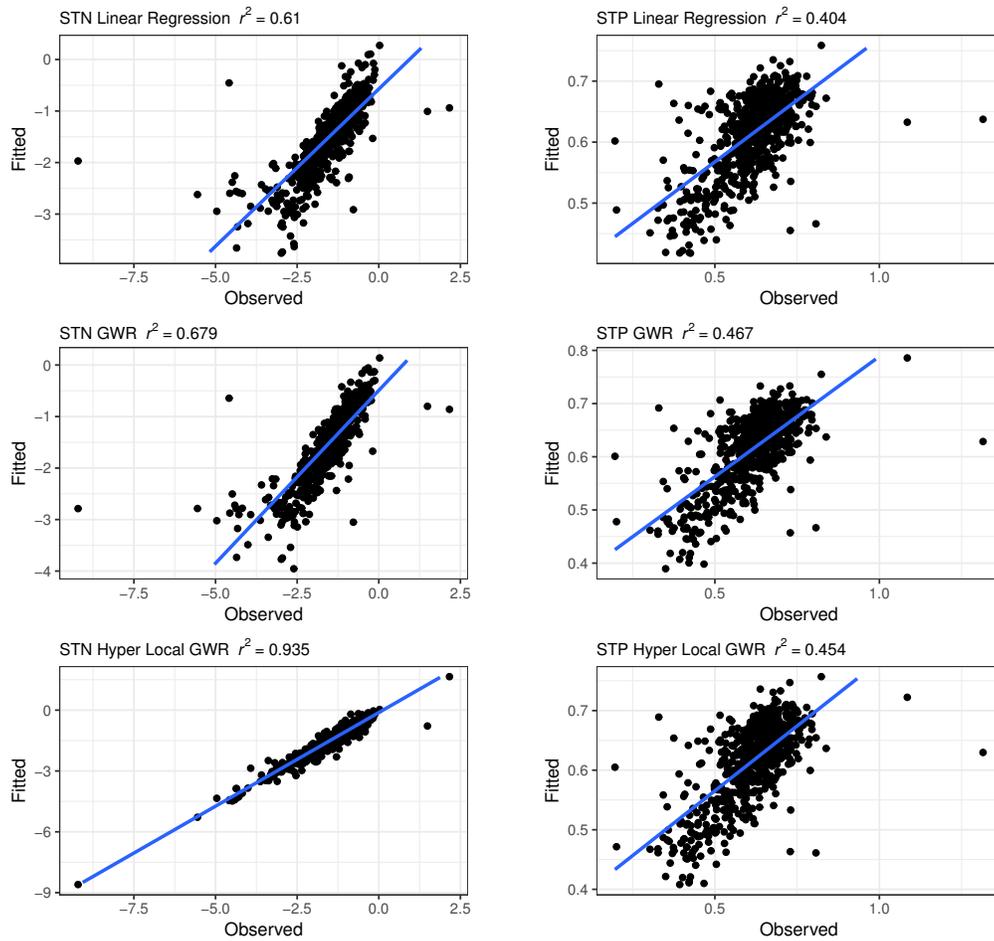


Figure 7: Fitted values for STN and STP arising from linear regression, standard GWR and hyper local GWR models against observed values.

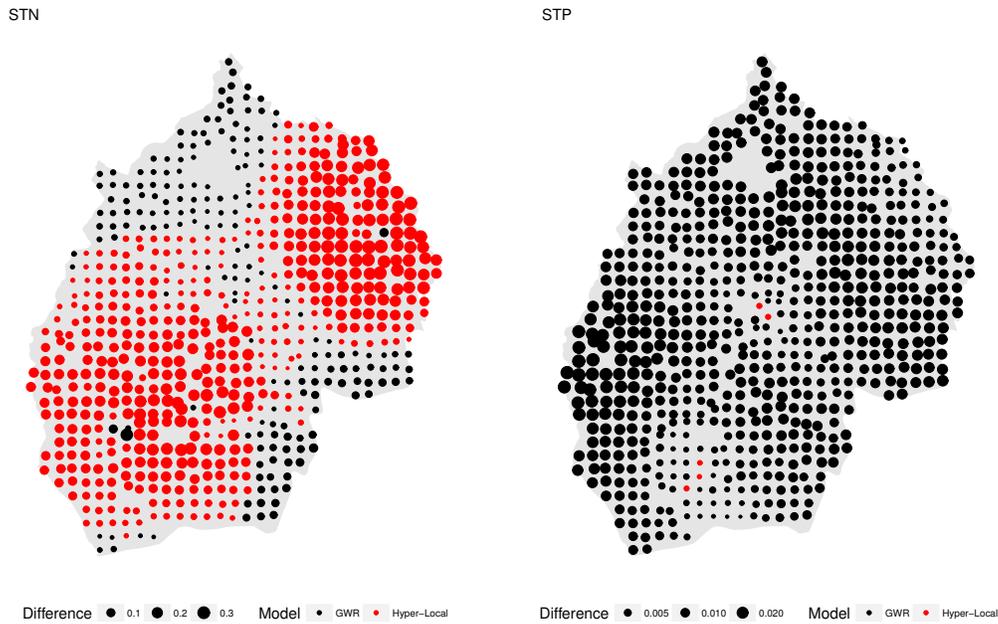


Figure 8: Maps of the difference in local R^2 values under standard GWR and hyper-local GWR models. Locations in red indicate where a hyper-local GWR model resulted in a better fitting model and (black for GWR) and the size of the plot characters indicate the magnitude of the difference.

well-informed by the six covariates. For STP, hyper-local GWR suggests more moderately spatially-varying relationships but where local model fits are similar (slightly weaker) to that found for standard GWR. Thus, the application of hyper-local GWR provides little value to an extended use of nearby data points with often fewer covariates for its local regressions. The STP process, is in general, not well-informed by the six covariates.

4 Discussion

GWR is an inherently exploratory approach for examining and investigating process non-stationarity in data relationships. The hyper-local GWR extends these investigations further. Whereas a standard GWR employs a one-size-fits-all bandwidth and a one-size-fits-all local regression model, the hyper-local GWR approach evaluates different kernel bandwidths and models at each location. It provides an alternative and complementary interpretation of localized regression by locally selecting the most parsimonious model (by local sample and covariate size), for which spatially distributed coefficient estimates and t -values can also be found. The local selection of the most parsimonious model is analogous to what is commonly done in a global analysis, where a summary of the full model is presented alongside a reduced, selected covariates model.

The investigations show that where the non-stationarity of relationships tend towards the global, as with STP, the results are similar to a standard GWR (compare Figures 3 and 6). However, where localized spatial heterogeneity and spatial non-stationarity *are* present, as with STN, the hyper-local GWR provides a more spatially nuanced indication of the localization than a standard GWR analysis (compare Figures 2 and 5). Thus the hyper-local GWR results can be used to guide the direction of the next analytical steps. Further analysis of the STN could consider adopting a more sophisticated spatially-varying coefficient model (e.g., [18]), including models that accounts for non-linearity (e.g., [2]). Further analysis of STP could consider a spatially-autocorrelated regression given that its GWR analyses were not entirely promising (e.g., [21]).

Determining local bandwidth size and local covariate selection is also in the same spirit as (but with entirely different objectives to) the GWR models of Paez et al. [34,35], Wheeler [42], and Yoneoka et al. [43]. These are analogous to developments in local (attribute-space) regression [29,40] from which GWR originates [5,31], but the GWR models of Paez only do local bandwidths (not local covariate selection) and the GWR models of Wheeler and Yoneoka only do local covariate selection (not local bandwidths). The exploratory and enhanced spatial nuance of hyper-local GWR reflects recent developments within the broad family of GWR methods that has promoted wider consideration of scale and distance. These include hierarchical GWR models [25], consideration of distance metrics [10,30], and flexible bandwidth GWR models [17,26] that select different bandwidths for each dependent/independent data relationship, rather than for each location as here. These multi-scale GWR models are closely aligned to the spatially-varying coefficient models of Gelfand et al. [18] and Murakami et al. [32].

There is a computational cost to hyper-local GWR approaches which evaluate a model for each bandwidth at each location, rather than a standard GWR, which just evaluates a single bandwidth and a single model. In terms of computing time, a standard GWR using the GWmodel package v2.0-5 [19] in R took 2.96 seconds to run on this data. The hyper-local approach took longer—20.7 minutes—because of the number of calculations but also because the algorithm has been transparently (rather than efficiently) coded. The data and code used in this analysis have been made available (see the acknowledgments section).

There are a number of considerations relating to the GWR models applied and demonstrated in this study. The first is collinearity. Standard GWR and hyper-local GWR are not designed to address collinearity issues, but here hyper-local GWR could be adapted to mitigate against such issues, in a similar manner to that proposed for standard GWR (e.g., [2,3,8]). In the hyper-local GWR approach described here, local models were selected under different bandwidths (i.e., with different local data subsets) at each location. Such model selection procedures identify the most parsimonious model and implicitly tend to select explanatory variables that are not collinear. The second concerns multiple hypothesis tests (MHTs) and spatial heteroskedasticity in the error term. Presenting GWR *t*-values in an uncorrected form can lead to the false discovery rate problem. Here the MHT corrections suggested by da Silva and Fotheringham [13] could be adopted for both standard and hyper-local GWR *t*-value outputs. Similarly GWR models that account for spatial heteroskedasticity in the error term can be found in Paez et al. [34,35] and in Fotheringham et al. [16] and Shen et al. [37], and the latter provides a procedure to similarly adapt a hyper-local GWR model. However, such GWR models have not been widely adopted, mainly due to inherent inferential issues, and as such are only viewed as exploratory (e.g., see Shen et al. [37]). A third consideration is the choice of kernel, the bandwidth type (fixed by dis-

tance, as used here, or fixed by sample size), and the choice of distance metric, all of which effect perspectives of coefficient non-stationarity. It would be interesting to examine the degree of difference between the STN and STP standard and hyper-local GWR models under such different parameterization choices. Gollini et al. [19] provides overviews of these considerations. Fourth, another area of future work is to examine the potential for overfitting with hyper-local methods. One way to test whether hyper-local GWR describes the study data significantly better than standard GWR would be to adapt the F -test procedure given in Leung et al. [27]. Here, instead of assessing standard GWR against the global linear regression, hyper-local GWR is assessed against standard GWR, where the null hypothesis is no significant difference between models. A fifth and more salient consideration for the research described in this paper is the use of AIC scores to select both local bandwidths and local regression models. AIC [1, 25] seeks to optimize model parsimony by trading off prediction accuracy and complexity. Other measures of fit could be applied including some kind of cross-validation measure of residual errors. There have been a number of arguments made in the context of information theory about the choice of model selection method and their associated measures of fit, and Li and Lam [28] review variable selection methods in GWR frameworks. They compared Step-AIC in GWR, GWR-Lasso, and GWR-Ridge models, noting that they are a function of zero-power, one-power, and two-power, respectively, of the explanatory variables in the models. This essentially frames the relationships between model selection within the elastic net. In terms of information criteria, alternatives to AIC exist such as Bayesian Information Criterion (BIC) and Deviance Information Criterion (DIC) [39]. Future work will investigate these and CV approaches as they would be expected to result in different local model selection. The key in determining which model selection method to use is to understand the logics of each approach and how they relate to the study objectives and even the underlying objectives of data collection. For example, AIC and BIC provide different approaches for model comparison [7]. BIC seeks to determine the “true” model and, if any particular candidate model represents the genuine data-generating mechanism, BIC will select such a model. It is said to be asymptotically consistent because it seeks to select the true model. By contrast AIC seeks to pragmatically select a model by trading-off explanations of the data with prediction strength. Despite these theoretical differences, Spiegelhalter et al. [38] note that “it is perhaps therefore rather surprising how often these two criteria produce similar rankings of candidate models” (p. 486) with the only real differences found in the size of the penalty scores [14]. Future work for both hyper-local and standard GWR will investigate the use of different model selection criteria, the logics associated with the local models being constructed and the underlying process spatial heterogeneity.

5 Conclusions

Local statistical approaches such as GWR are inherently exploratory in nature. They seek to confirm or refute spatial heterogeneity in spatial data structure, processes and statistical relationships. The *hyper-local* GWR approach described in this paper provides a useful counter view of local regression modeling to that found with standard GWR. The results of this study show that a standard GWR analysis may underestimate the degree of localized spatial heterogeneity in data relationships where it is strongly present (as in the STN case study) and may overestimate it where spatial homogeneity is present (as in the STP case

study). Standard GWR applies the same regression model at each location and uniformly sets the same kernel bandwidth everywhere. The hyper-local GWR approach evaluates different kernel bandwidths at each location and selects the most parsimonious local regression model. Where spatial non-stationarity exists, the hyper-local GWR provides a more spatially nuanced indication of the localization than a standard GWR analysis and can be used to suggest the direction of further analyses and investigations. Undertaking a hyper-local GWR alongside a standard GWR allows coefficient estimates, t -values and bandwidths to be compared for differences and similarities. Specifically, a dual GWR approach that examines the spatial distribution of local covariate selection and the local bandwidth size supports a deeper understanding of the local and scale-related characteristics of the spatial process under investigation.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) and the Natural Environment Research Council (NERC) Newton Fund through the China-UK collaborative research on critical zone science (No. 41571130083 and NE/N007433/1), the NSFC (No.41530854) and UK Biotechnology and Biological Sciences Research Council grant (BBSRC BB/J004308/1). All of the analyses and mapping were undertaken in R 3.4.3 the open source statistical software. The GWR analyses used the GWmodel package, v2.0-1 [19]. The data and code used to generate the analysis and all of the figures and tables can be accessed from <https://github.com/lexcomber/hyperlocalGWR>. AC conceived, and led the design and implementation of the analysis; YW collected and processed the data used in the analysis; PH guided the analysis. All authors contributed to interpreting the results, refining the analysis, and the writing of the manuscript.

References

- [1] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd Symposium on Information Theory, BN Petrov, F Csaki (eds.)*, (1973), Akademiai Kiado, Budapest, pp. 267–281.
- [2] BÁRCENA, M., MENÉNDEZ, P., PALACIOS, M., AND TUSELL, F. Alleviating the effect of collinearity in geographically weighted regression. *Journal of Geographical Systems* 16, 4 (2014), 441–466. doi:10.1007/s10109-014-0199-6.
- [3] BASILE, R., DURBÁN, M., MÍNGUEZ, R., MONTERO, J. M., AND MUR, J. Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control* 48 (2014), 229–245. doi:10.1016/j.jedc.2014.06.011.
- [4] BRUNSDON, C., FOTHERINGHAM, A., AND CHARLTON, M. Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems* 26, 6 (2002), 501–524. doi:10.1016/S0198-9715(01)00009-6.

- [5] BRUNSDON, C., FOTHERINGHAM, A. S., AND CHARLTON, M. E. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis* 28, 4 (1996), 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x.
- [6] BRUNSDON, C., FOTHERINGHAM, S., AND CHARLTON, M. Geographically weighted discriminant analysis. *Geographical Analysis* 39, 4 (2007), 376–396. doi:10.1111/j.1538-4632.2007.00709.x.
- [7] BURNHAM, K. P., AND ANDERSON, D. R. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research* 33, 2 (2004), 261–304. doi:10.1177/0049124104268644.
- [8] CHILÈS, J., AND DELFINER, P. *Geostatistics: Modeling spatial uncertainty*. John Wiley & Sons, New York., 1999.
- [9] COMBER, A., BRUNSDON, C., CHARLTON, M., AND HARRIS, P. Geographically weighted correspondence matrices for local error reporting and change analyses: mapping the spatial distribution of errors and change. *Remote Sensing Letters* 8, 3 (2017), 234–243. doi:10.1080/2150704X.2016.1258126.
- [10] COMBER, A., CHI, K., QUANG HUY, M., NGUYEN, Q., LU, B., HUU PHE, H., AND HARRIS, P. Distance metric choice can both reduce and induce collinearity in geographically weighted regression. *Environment and Planning B: Urban Analytics and City Science* (2018). doi:10.1177/2399808318784017.
- [11] COMBER, A., LI, T., LÜ, Y., FU, B., AND HARRIS, P. Geographically weighted structural equation models: spatial variation in the drivers of environmental restoration effectiveness. In *Societal Geo-Innovation*. 20th AGILE Conference Proceedings, 2017.
- [12] COMBER, A. J. Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. *Remote Sensing Letters* 4, 4 (2013), 373–380. doi:10.1080/2150704X.2012.736694.
- [13] DA SILVA, A. R., AND FOTHERINGHAM, A. S. The multiple testing issue in geographically weighted regression. *Geographical Analysis* 48, 3 (2016), 233–247. doi:10.1111/gean.12084.
- [14] DZIAK, J. J., COFFMAN, D. L., LANZA, S. T., AND LI, R. Sensitivity and specificity of information criteria. *PeerJ PrePrints* (2017). doi:10.7287/peerj.preprints.1103v3.
- [15] FOODY, G. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing* 26, 6 (2005), 1217–1228. doi:10.1080/01431160512331326521.
- [16] FOTHERINGHAM, A. S., BRUNSDON, C., AND CHARLTON, M. *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. Wiley, Chichester, 2002.
- [17] FOTHERINGHAM, A. S., YANG, W., AND KANG, W. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers* 107, 6 (2017), 1247–1265. doi:10.1080/24694452.2017.1352480.

- [18] GELFAND, A. E., KIM, H.-J., SIRMANS, C., AND BANERJEE, S. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98, 462 (2003), 387–396. doi:10.1198/016214503000170.
- [19] GOLLINI, I., LU, B., CHARLTON, M., BRUNSDON, C., AND HARRIS, P. Gwmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software* 63, 17 (2015), 1–50. doi:10.18637/jss.v063.i17.
- [20] HARRIS, P., BRUNSDON, C., AND CHARLTON, M. Geographically weighted principal components analysis. *International Journal of Geographical Information Science* 25, 10 (2011), 1717–1736. doi:10.1080/13658816.2011.554838.
- [21] HARRIS, P., BRUNSDON, C., LU, B., NAKAYA, T., AND CHARLTON, M. Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models. *Spatial Statistics* 21 (2017), 241–261. doi:10.1016/j.spasta.2017.07.006.
- [22] HARRIS, P., CHARLTON, M., AND FOTHERINGHAM, A. S. Moving window kriging with geographically weighted variograms. *Stochastic Environmental Research and Risk Assessment* 24, 8 (2010), 1193–1209.
- [23] HARRIS, P., FOTHERINGHAM, A., CRESPO, R., AND CHARLTON, M. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences* 42, 6 (2010), 657–680. doi:10.1007/s11004-010-9284-7.
- [24] HARRIS, P., FOTHERINGHAM, A. S., AND J. S. Robust geographically weighted regression: a technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. *Annals of the Association of American Geographers* 100, 2 (2010), 286–306. doi:10.1080/00045600903550378.
- [25] HURVICH, C. M., SIMONOFF, J. S., AND TSAI, C.-L. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60, 2 (1998), 271–293. doi:10.1111/1467-9868.00125.
- [26] LEONG, Y.-Y., AND YUE, J. C. A modification to geographically weighted regression. *International journal of health geographics* 16, 1 (2017), 11. doi:10.1186/s12942-017-0085-9.
- [27] LEUNG, Y., MEI, C.-L., AND ZHANG, W.-X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A* 32, 1 (2000), 9–32. doi:10.1068/a3162.
- [28] LI, K., AND LAM, N. S. Geographically weighted elastic net: A variable-selection and modeling method under the spatially nonstationary condition. *Annals of the American Association of Geographers* (2018), 1–19. doi:10.1080/24694452.2018.1425129.
- [29] LOADER, C. Smoothing: local regression techniques. In *Handbook of Computational Statistics*. Springer, 2012, pp. 571–596. doi:10.1007/978-3-642-21551-3_20.
- [30] LU, B., BRUNSDON, C., CHARLTON, M., AND HARRIS, P. Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science* 31, 5 (2017), 982–998. doi:10.1080/13658816.2016.1263731.

- [31] MCMILLEN, D. P., AND MCDONALD, J. F. A nonparametric analysis of employment density in a polycentric city. *Journal of Regional Science* 37, 4 (1997), 591–612. doi:10.1111/0022-4146.00071.
- [32] MURAKAMI, D., YOSHIDA, T., SEYA, H., GRIFFITH, D. A., AND YAMAGATA, Y. A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics* 19 (2017), 68–89. doi:10.1016/j.spasta.2016.12.001.
- [33] OPENSHAW, S. Developing GIS-relevant zone-based spatial analysis methods. *Spatial analysis: modelling in a GIS environment* (1996), 55–73.
- [34] PÁEZ, A., UCHIDA, T., AND MIYAMOTO, K. A general framework for estimation and inference of geographically weighted regression models: 1. location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A* 34, 4 (2002), 733–754. doi:10.1068/a34110.
- [35] PÁEZ, A., UCHIDA, T., AND MIYAMOTO, K. A general framework for estimation and inference of geographically weighted regression models: 2. spatial association and model specification tests. *Environment and Planning A* 34, 5 (2002), 883–904. doi:10.1068/a34133.
- [36] RIPLEY, B., VENABLES, B., BATES, D. M., HORNIK, K., GEBHARDT, A., FIRTH, D., AND RIPLEY, M. B. Package ‘mass’. *Cran R* (2013).
- [37] SHEN, S.-L., MEI, C.-L., AND ZHANG, Y.-J. Spatially varying coefficient models: testing for spatial heteroscedasticity and reweighting estimation of the coefficients. *Environment and Planning A* 43, 7 (2011), 1723–1745. doi:10.1068/a43201.
- [38] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., AND LINDE, A. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76, 3 (2014), 485–493. doi:10.1111/rssb.12062.
- [39] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., AND VAN DER LINDE, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 4 (2002), 583–639. doi:10.1111/1467-9868.00353.
- [40] VIDAURRE, D., BIELZA, C., AND LARRAÑAGA, P. Lazy lasso for local regression. *Computational Statistics* 27, 3 (2012), 531–550. doi:10.1007/s00180-011-0274-0.
- [41] WANG, Y., ZHANG, X., AND HUANG, C. Spatial variability of soil total nitrogen and soil total phosphorus under different land uses in a small watershed on the Loess Plateau, China. *Geoderma* 150, 1-2 (2009), 141–149. doi:10.1016/j.geoderma.2009.01.021.
- [42] WHEELER, D. C. Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and planning A* 41, 3 (2009), 722–742. doi:10.1068/a40256.
- [43] YONEOKA, D., SAITO, E., AND NAKAOKA, S. New algorithm for constructing area-based index with geographical heterogeneities and variable selection: An application to gastric cancer screening. *Scientific reports* 6, 26582 (2016), 1–7. doi:10.1038/srep26582.