Spatial considerations in the analysis of conservation data: evaluating the presence of *Acer campestre* (field maple) – space is special

Alexis Comber[1], Steve Carver[1], Carol Ximena Garzon-Lopezˆ2, Paul Harris[3], Duccio Rocchini[2]

[1] School of Geography, University of Leeds, Leeds, LS2 9JT, UK

Email: {a.comber; s.j.carver} @leeds.ac.uk

[2] Fondazione Edmund Mach, 38010 S. Michele all'Adige, Italy

Email: duccio.rocchini@fmach.it

[3] Rothamsted Research, North Wyke, EX20 2SB, UK

Email: paul.harris@rothamsted.ac.uk

Abstract

TBC

1. Introduction

This paper emphasises the need for *Geography Googles* in applied geographical analyses of data that increasingly have a spatial comonent in the form of location, for instance form GPS. It argues that *space is special* and encourages explicitly spatial ways of thinking. These ideas are posited in the context of nearly all data being spatial (i.e location is included), and the need for slightly more informed approaches for analysing space and location than provided by standard statistical models.

Blah blah blah (Lex to complete but any suggestions welcome)

In order to demenstrate the value of spatial dependence emethods, this paper tests the model suggested by Coudun et al (2006) but does this using a Geographically Weighted (GW) framework (Brunsdon et al, 1996) in this case a GW Regression (GWR). Coudun et al (2006) found the presence of *Acer campestre* in France to be significantly related to two factors rainfall and evapotranspiration. Here Coudon's analysis is extended to account for spatial autocorrelation in the variables which occurs when changes in properties of nearby features are found to be correlated and this contradicts the underlying assumption of independence in statistical analysis and inference. The result is spatial non-stationarity when the statistical pattern or relationship observed in one location differs from that in another. To explore these, this paper applies a GWR analysis to examine the spatial variation in the relationships between the predictor variables and Acer campestre presence to test for the presence of local, non-stationary relationships.

2. Background

Understanding species distribution and spatial dependencies is a key concept in ecological research (**REF needed**), in particular when dealing with biodiversity monitoring for conservation practices (**REF needed**). Several papers have dealt with the importance of explicitly taking space into account when trying to model the distribution of a certain species, and, overall, its change in space and time (**CITATIONS here**). The description of the distribution of biodiversity at different spatial and temporal scales has long been the focus of ecology and biogeography. Reliable descriptions of species distributions are fundamental for conservation and research purposes (Dormann, 2007). However, when modelling species spread, a number of statistical problems must be solved before spatial models can be built such as: the uncertainty related to the sampling of the modelled species (Rocchini et al., 2011), the correlation among predictors (**Dormann et al....**), the potential variability over space of the predictors being used (Rocchin et al., 2016, CAGEO), the spatio-temporal dependency of the predictors (Zuur and Ieno, 2016). Among these, spatial non stationairty and scale dependence are expected to impact most of the ecological processes and relative patterns such as the distribution of biodiversity over space (Foody, 2004).

The importance of spatial considerations has long been recognised in a number of disciplines: Fischer & Gosset (1935) in crop science; Kolmogorov (1941), Gandin (1965) in meteorology; Krige (1951) and Matheron (1963) in mining; Matérn (1960) in forestry; theoretical developments by Moran (1950) and Yaglom (1955); Berry and Marble (1968) and Chorley and Haggett (1967) in geography; and Legendre and Legendre (1991) in ecology.

The major issue with modelling spatial data is the lack of independence of the spatial objects and the assumption of stationarity in the processes being modelled. These are particularly true with socio-economic data, but are also true of environmental and ecological data. Both concerns relate to Tobler's dictum (the first Law of Geography) that: '*everything is related to everything else, but near things are more related than distant things*' (Tobler 1970). However, in spite of numerous methodological advances in addressing such spatial modelling problems, policy related research is still routinely informed by non-spatial modelling practices, which are heroic at best, and ill-advised at worst. The lineage, from the late 1970s onwards, of such methodological advances can be loosely traced through spatial statistics texts from Journel and Huigbrechts (1978) to Cressie (1993) to Chilès and Delfiner (1999) to Cressie and Wilke (2011).

Focusing on regression models, the main drawback in assuming observations are independent is that any spatial dependencies turn up in the residuals, as they are not explicitly added to the model. For area-based data, ways to account for this include Besag's (1974) conditionally autoregressive model or Anselin's (1988) spatially autoregressive model. In both cases, a criterion is needed for determining 'nearby' and is often supplied by the entries in a matrix of adjacencies. Wall (2004) notes that spatial covariance structures are implied by such matrices, which sometimes act as confounders in the analysis implying counterintuitive spatial processes. Vastly different covariance structures can be observed for different values of the autoregressive parameter. This is worrying for area-based modelling and is similar in nature to that found in Openshaw and Taylor's (1979) analysis of voting data in Iowa US, where different arrangements of the spatial units yielded correlations from -1 to +1 for the same pre-aggregated data. For point-based data, geostatistical regressions are commonly used, where spatial dependencies are modelled directly via the variogram, a function specifying a deterministic relationship between point pairs and their correlation. Recent advances have also seen geostatistical models applied in health studies where area-based data is common (Goovaerts 2009). These models are worthy in that they do not suffer from the pitfalls described by Wall (2004).

Spatial information can also be accounted for by expressing the regression's coefficients as functions of the spatial coordinates (Casetti 1982; Gorr and Oligschaeger 1994), permitting a model of relationship nonstationarity. A major advance in this area was advent of the GWR model (Brunsdon et al. 1996). Here spatial structures in the data are incorporated into the model through a continuous distance-decay weighting scheme. The spatial extent of the scheme is controlled through the bandwidth parameter, which is optimised to give the best model fit to the data. The resultant localised regression coefficients are mapped to display their spatial variation. Brunsdon et al. (1998) and Harris et al. (2015) provide some simple inferential mechanisms to determine whether the coefficients exhibit significant spatial variation; and other notable advances in the GWR method can be found in Nakaya et al. (2005), Wheeler (2007), Huang et al. (2010), Harris et al. (2011), Brunsdon et al. (2012) and Silva and Fotheringham (2015). However, a fully-coherent inferential structure of GWR remains to be developed. It may be that, given GWR's kernel smoothing roots (e.g. see Cleveland 1979), the technique is best reserved for use as an exploratory and highly visual tool only, to act as a valuable precursor to a more sophisticated spatially-varying coefficient model, such as those proposed by Gelfand et al. (2003) or Assunção (2003).

The over-riding issue is that dependence and some form of nonstationarity are endemic in spatial data, and model forms should be made widely available which allow users to explore these structures both pre- and post-modelling. When spatial non-stationarity holds, the predictor and response variables is expected to change over space in an area. Under spatial non-stationarity the relationship between the distribution of a certain species and the predictors shaping it may change over space. Thus any modelling activity should also incorporate mechanisms to handle these characteristics of spatial data. This is an important challenge and this paper focuses on the modelling of spatial nonstationarity via the GWR model in order to demonstrate how to address process heterogeneity in relationships. The GWR model itself is formally described in Section 3, both in a basic form and an adapted form to deal with potential problems of localised collinearity in the predictor variables (Brunsdon et al. 2012; Gollini et al. 2015).

3. Methods

3.1 Data and Study Area

Data recorded between 1980 and 2010 describing the presence of *Acer campestre* was downloaded from GBIF using the `dismo` R package, and subsetted for the UK. The data contained 22,701 records whose spatial distribution are shown in Figure 1, with a median of 551 records per year, a 1st quartile of 372 and a 732.3 quartile of 917 records. The data for all years were summed over Ordnance Survey 10km grids because of the uneven distribution in time and space. Potential alternative approaches including generating absence points, background data (Phillips et al. 2009) to charcaterise study area environments or pseudo-absences (eg VanDerWal et al., 2009), indicating where absences might occur. However, pseudo absence approaches require a number of assumptions and lack statistical methods for handling the overlap between presence and background points (Ward et al. 2009; Phillips and Elith, 2011), absence data may be biased and or incomplete (Kery et al., 2010) and background data approaches generate the same measures irrespective of where the species is observed (Hijmans and Elith, 2015).

The study by Coudun et al (2006) found the presence of *Acer campestre* to be significantly related to Autumn rainfall and actual Thornthwaite evapotranspiration. In this study rainfall, Potential evapotranspiration (PET) and wildness index data were used to construct a series of models for predicting the density of *Acer campestre* occurance. Data on rainfall were downloaded from the NERC Environmental Information Data Centre (Tanguy et al, 2015) which provides monthly 1km estimated rainfall data for each year. The average Autumn (3 month) rainfall was calculated for each 1km. Mean annual PET data were downloaded from the CGIAR Consortium for Spatial Information (Trabucco and Zomer, 2009) which provides global data at textbf{0.0083 degrees, approximately 1km}. Finally, a wilder dataset was included in the model. This was to explore the degree to the presence of *Acer campestre* may be related to anthropogenic disturbance – anecdotally this species is frequently used as

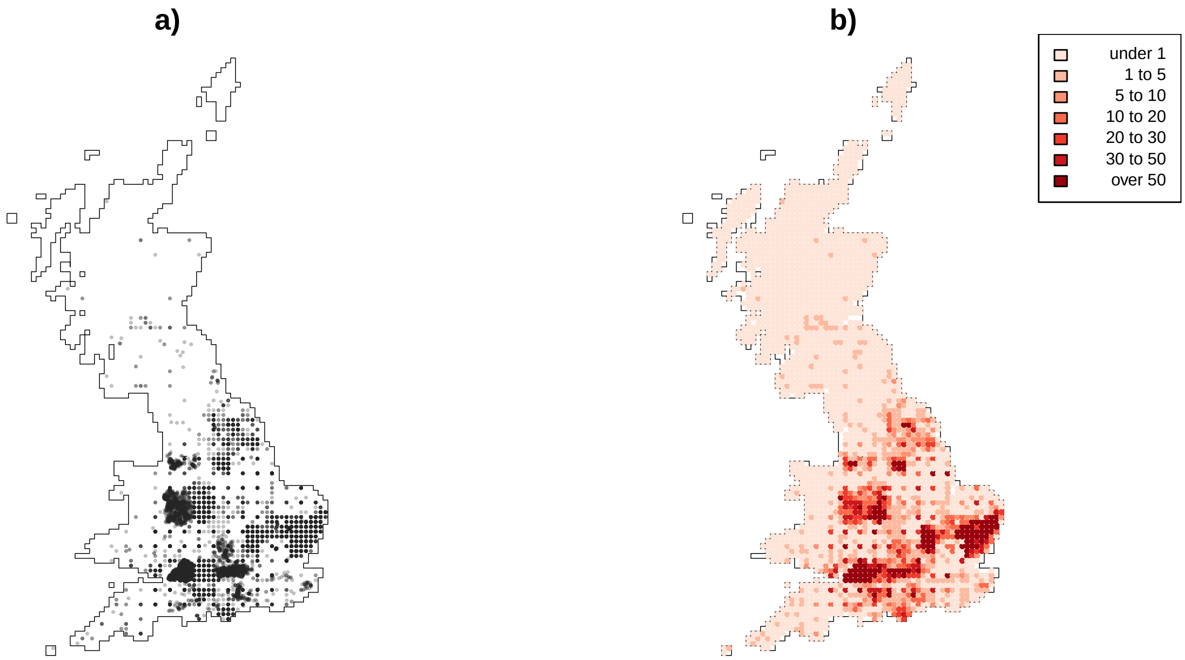

Figure 1: a) The raw data points with a transparency term to show density of points, and b) Data summed over OS 10km grid cells.

an ornamental tree and found in field margins and hedges. The Wilderness Quality Index data were generated for the whole of Europe at 1km resolution as described in Kuiters et al (2011)

. . . .more from Steve

WQI can be considered as measure of non-anthropogenic activity. Each of these datasets were spatially aggregated over the OS 10m grid cells to generate mean values as shown in Figure 2.

3.2 Analysis

A multi-stage analysis was applied to model *Acer campestre* distributions. First, exploratory analyses were undertaken using a standard OLS regression to model distributions as an initial step. This identified significant predictor variables, under the assumption that relationships between predictor variables (rainfall, PET and wilderness) and species distributions are stationary (i.e. global). Then a GWR analysis was applied to examine the spatial variation in the relationships between the predictor variables and *Acer campestre* distributions (i.e. to test for the presence of local, non-stationary relationships). In overview, GW approaches use a moving window or kernel that passes through the study area. At each location being considered, data under the window are used to make a local calculation of some kind, such as a regression. The data are weighted by their distance to the kernel centre and in this way GW approaches construct a series of models at discrete locations in the study area. This is in contrast to global models, that consider all of the data (usually) in a single analysis of all data in the study area.

Next, the presence of local collinearity amongst the predictor variables was tested. This is a critical step in any GWR analysis but one that is usually overlooked. Collinearity occurs when variables exhibit linear or near linear relationships. Strong collinearity will affect model reliability and precision, generate unstable parameter estimates, inflated standard errors and inferential biases (Dormann et al 2013), and there may be problems in separating variable effects (Meloun et al. 2002). In a GWR analyses, collinearity may occur locally, with the construction of localised regressions, even when it

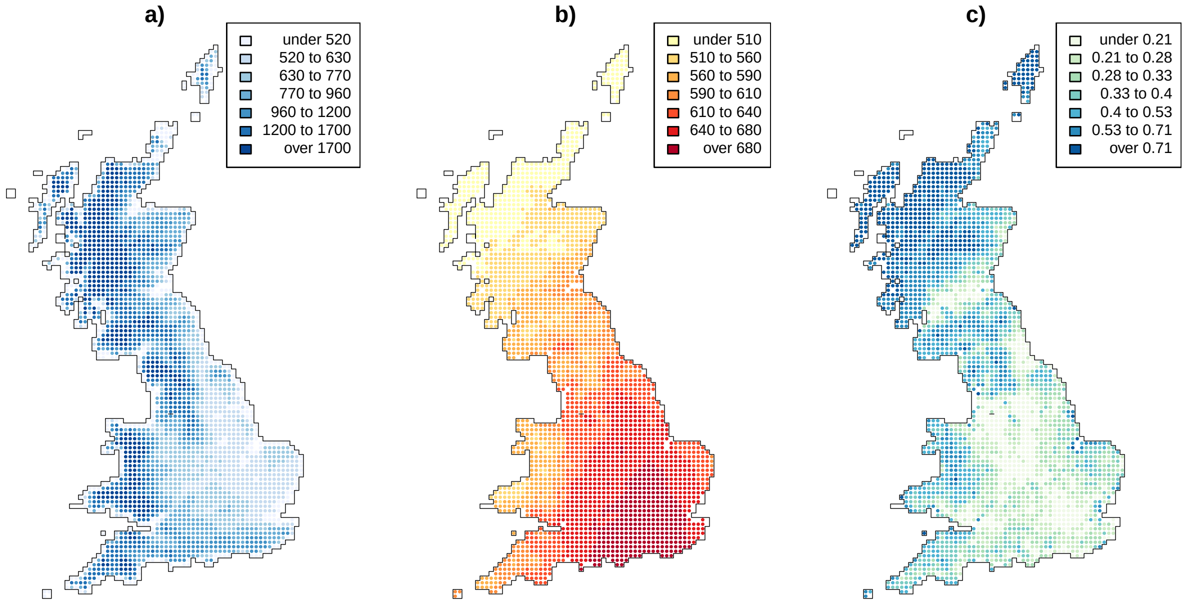

Figure 2: The data used to construct the species model, a) mean monthly Autumn Rain (mm), b) mean annual potential evapotranspiration (PET), and c) Mean Wildness Quality Index.

is not observed globally (Wheeler and Tiefelsdorf, 2005; Wheeler 2007, 2009, 2013; Brunsdon et al 2012). A number of approaches exist to address collinearity in regression modelling, such as partial least squares regression, principal component analysis regression and ridge regression (Hoerl 1962; Hoerl and Kennard 1970). Ridge regression is a penalised model where extensions, such as the lasso and the elastic net also provide predictor variable sub-set selection (e.g. Zou and Hastie 2005). All such models could be adapted to a localised form and Wheeler (2007; 2009) has proposed both ridge and lasso versions of GWR to address any detrimental local collinearity effects. A related, locally-compensated ridge GWR model is detailed in Brunsdon et al (2012), Lu et al. (2014) and Gollini et al (2015). It has advantages over the ridge GWR model of Wheeler (2007) in that a ridge term is applied locally and not globally. These studies also describe associated local collinearity diagnostics for GWR, such as the use of local correlations amongst pairs of predictors, local Variance Inflation Factors (VIFs) for each predictor, local variance decomposition proportions (VDPs) and the local condition numbers (CNs). The key point about locally-compensated ridge GWR, is that a local ridge term is only applied where it is needed – when the local CN is above a pre-specified value in this case 30, which is a standard heuristic.

This study undertakes a GWR analysis, with a locally-compensated ridge term if necessary, over a 200m grid of points covering the study area, with the aim of examining the spatial distribution of coefficient estimates predicting house price and their spatial variation. Euclidean distances were used to weight data points under the kernel. These distances better reflect the spatial processes and relationships in environmental systems than network distance (Comber et al., 2008). For the kernel, an adaptive bi-square weighting function was applied, although a number of kernel functions can be specified for GW models as discussed in Gollini et al (2015). This generates higher weights at locations very near to the kernel centre relative to those towards the edge. For each data point (P_j) under the kernel (with a given bandwidth), a weight $w_{i,j}$ is calculated based on its distance to the centre of the kernel (K_i) as follows:

$$w_{i,j} = 1 - ((d_{i,j})^2/b^2) \tag{1}$$

where $d_{i,j}$ is the distance in metres from the centre of the kernel K_i to the data point P_j and b is the bandwidth.

An optimum kernel bandwidth for GWR can be found by minimising a model fit diagnostic. Options include a leave-one-out cross-validation (CV) score (Bowman 1984; Brunsdon et al. 1996). This optimises model prediction accuracy and the Akaike Information Criterion (AIC) (Akaike 1973; Fotheringham et al. 2002) optimises model parsimony by trading off prediction accuracy and complexity. In this case, the CV approach was applied to specify all GWR models, all using the bi-square weighting kernel and distances between locations.

The standard GWR model is:

$$y_i = \beta_{i0} + \sum_{m}^{k=1} \beta_{ik} x_{ik} + \epsilon_i \tag{2}$$

where y_i is the response variable at location i, x_{ik} is the value of the kth predictor variable at location i, m is the number of predictor variables, β_{i0} is the intercept term at location i, β_{ik} is the local regression coefficient for the k^{th} predictor variable at location i and ϵ_i is the random error at location i. The result of the weighting means that data nearer to the kernel centre make a greater contribution to the estimation of local regression coefficients at each local regression calibration point i.

3.3 Code

All of the analyses and mappings were undertaken in R, the free open source statistical software. The Rmarkdown script used to produce this manuscript, including all the code used in the analysis and to produce the mapped figures, can be found at https://github.com/lexcomber/SpatEcolPap

4. Results

4.1 Exploratory Regressions

A standard OLS regression models was undertaken and the resultant coefficient estimates and significance values are shown in Table 1 below. PET and Wilderness were found to be significant predictors of Acer campestre distributions at the 5% level (i.e. with a less than 95% chance of occurring randomly). Interestingly, in contrast to the findings of Coudun et al (2006), mean Autumn rainfall was not found to be significantly associated with the *Acer campestre* distributions.

Table 1. The global regression co-efficient estimates.

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-28.124	13.218	-2.128	0.033
Mean annual PET	0.072	0.018	3.936	0.000
Mean Autumn rainfall	0.000	0.001	-0.233	0.815
Mean Wilderness Quality Index	-14.268	6.443	-2.215	0.027

Table 2. The variation of the coefficient estimates arising from a GWR analysis.

	Min	1st Qu	Median	3rd Qu	Max
Intercept	-20572.266	-27.726	0	13.433	37494.330
Mean annual PET	-51.335	-0.021	0	0.044	30.428
Mean Autumn rainfall	-3.086	0.000	0	0.002	3.372
Mean Wilderness Quality Index	-4841.516	-3.569	0	1.996	1391.310

The underlying theoretical framework provided by GWR tests for non-stationarity in processes and relationships between factors. A standard GWR analysis was undertaken and in this case an optimal adaptive bandwidth of 21 data points was determined using a cross-validation procedure.

The local coefficient estimates from this GWR model are shown in Table 2. They indicate considerable variation around the median in the degree to which increases in the predictor variables are associated with Acer campestre distributions. For example, considering the inter-quartile ranges shows that, in some places:

- An increase in PET of 100 values is associated with a decrease of -2.1 trees;
- That each increase of 0.3 in the wilderness index is associated with a decrease of 1 tree (-3.569 * 0.3); But in other locations:
- A decrease in PET of 100 values is associated with an increase of 4.4 trees;
- That each increase of 0.5 in the wilderness index is associated with an increase of 1 tree (1.996 * 0.5).

The local variation in coefficient estimates in Table 2 is in contrast to the global coefficient estimates in Table 1.

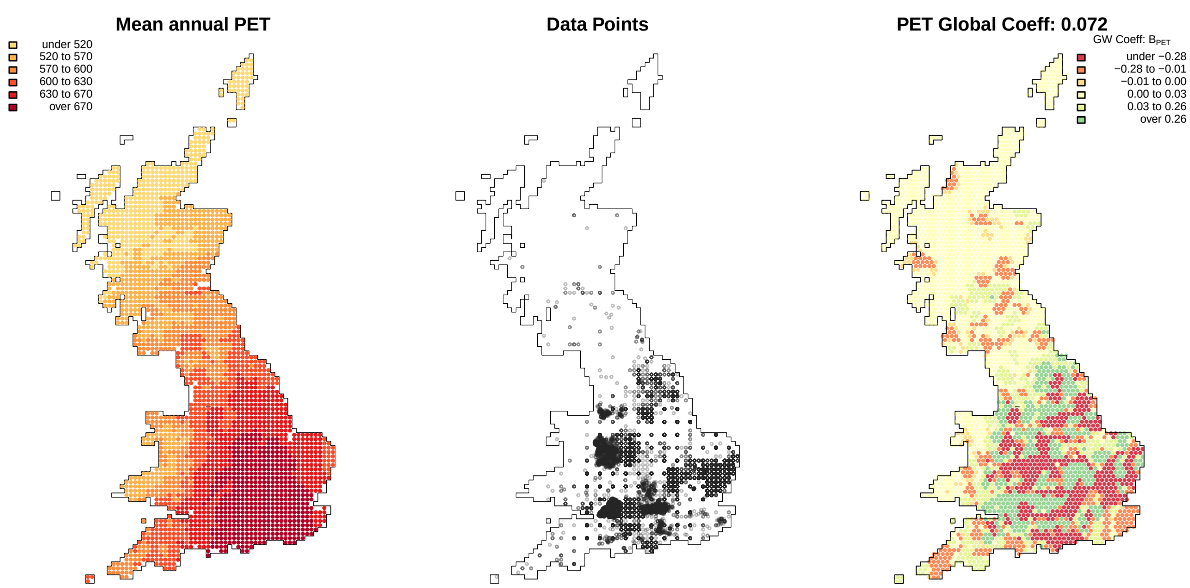

Figure 3: The spatial distribution of the mean annual PET coefficient estimates, with the context of the original PET and species data.

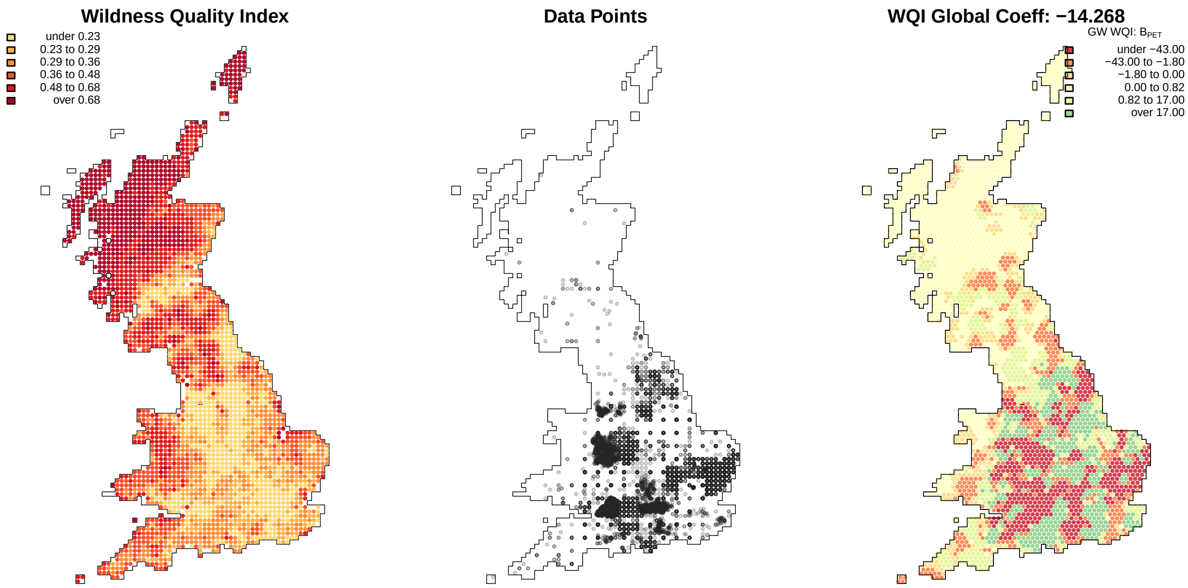

Figure 4: The spatial distribution of the Wilderness Quality Index coefficient estimates, with the context of the original WQI and species data.

4.2 GWR Local Collinearity Diagnostics

The potential for detrimental effects due to local collinearity has been ignored by nearly all of the GWR analyses reported in the literature, regardless of domain or subject. Collinearity occurs when one predictor variable has a strong positive or negative relationship with another, typically when it is less than -0.8 or greater than +0.8. Critically, collinearity may be absent when calculated globally (ie from all the data values), but may be present locally when a subset of the data is considered, as is the case with a GWR analysis. Table 3 shows the results of evaluating collinearity globally and locally using the GWR collinearity diagnostics tool included in the `GWmodel` R package.

Globally, the Variance Infation Factors (VIFs) are all less than 10, although 2 of the Variance Decomposition Proportions (VDPs) are greater than 0.5 and the Condition Number (CN) is greater than 30, using standard heuristics from Belsley et al (1980) and O'Brien (2007). Together these suggest the presence of variable collinearity when evaluated globally considering all data points together. Applying GWR collinearity diagnostics to the GWR model above generates local VIFs, VDPs and CNs at the same scale (i.e. using the same adaptive bandwidth of 21 data points). The results in Table 3 indicate a high degree of *local* collinearity in the CN, VIFs and VDPs. These values suggest that the application of a locally-compensated ridge GWR is warranted. The local collienarity measures are mapped in Figure 5.

Table 3. Global and local collinearity measures: Condition Numbers with Variance Infaltion Factors (VIFs) and Variance Decomposition Proportions (VDPs) for each predictor variable.

	Global	Local Min	Local 1st Qu	Local Median	Local 3rd Qu	Local Max
CN	43.899	93.767	408.129	571.023	822.609	5033.693
VIF PET	3.111	1.000	1.513	2.616	6.384	153.746
VIF Rainfall	1.189	1.000	1.515	2.594	5.216	156.274
VIF WQI	3.216	1.000	1.368	2.023	3.549	48.685

	Global	Local Min	Local 1st Qu	Local Median	Local 3rd Qu	Local Max
VDP PET	0.992	0.541	0.997	0.999	1.000	1.000
VDP Rainfall	0.007	0.000	0.119	0.392	0.673	0.998
VDP WQI	0.692	0.000	0.062	0.221	0.491	0.981

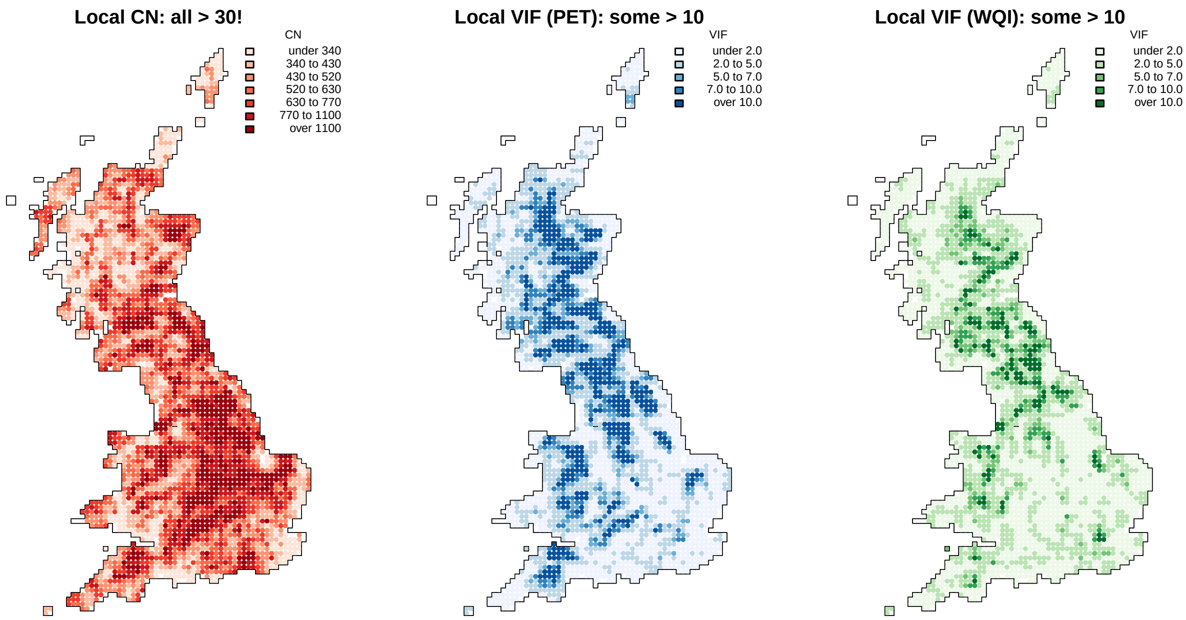

Figure 5: The spatial distribution of the local CNs and the Mean annual PET and Mean Wilderness Quality VIFs.

4.3 Final GW analysis

Having tested for and identified local collinearity, a locally compensated ridge GWR (LCR-GWR) was specified. This applies a GW regression but with a locally-compensated ridge term and fits local ridge regressions with their own ridge parameters (i.e., the ridge parameter varies across space), but only does this at locations where the local Condition Number is above a user-specified threshold. In this case the CN threshold was specified as 30. An optimal adaptive bandwidth of 21 data points was again determined using a cross-validation procedure. Figures 6 and 7 shows the spatial distribution of the original GWR coefficients, those determined using a locally compensated ridge GWR and a map of the differences between the two, for PET and for Wilderness Quality Index. In both cases there are large and potentially important differences between the coefficient estimates from the GWR and those from the LCR-GWR.

5. Discussion

In this paper a series of analyses were undertaken to demonstrate the application and value of explicitly spatial analyses, focusing on GWR. These develop local statistical models in order to test for spatial non-stationarity and are in contrast to standard, a-spatial, statistical approaches that assume the relationships between factors to be the same everywhere.

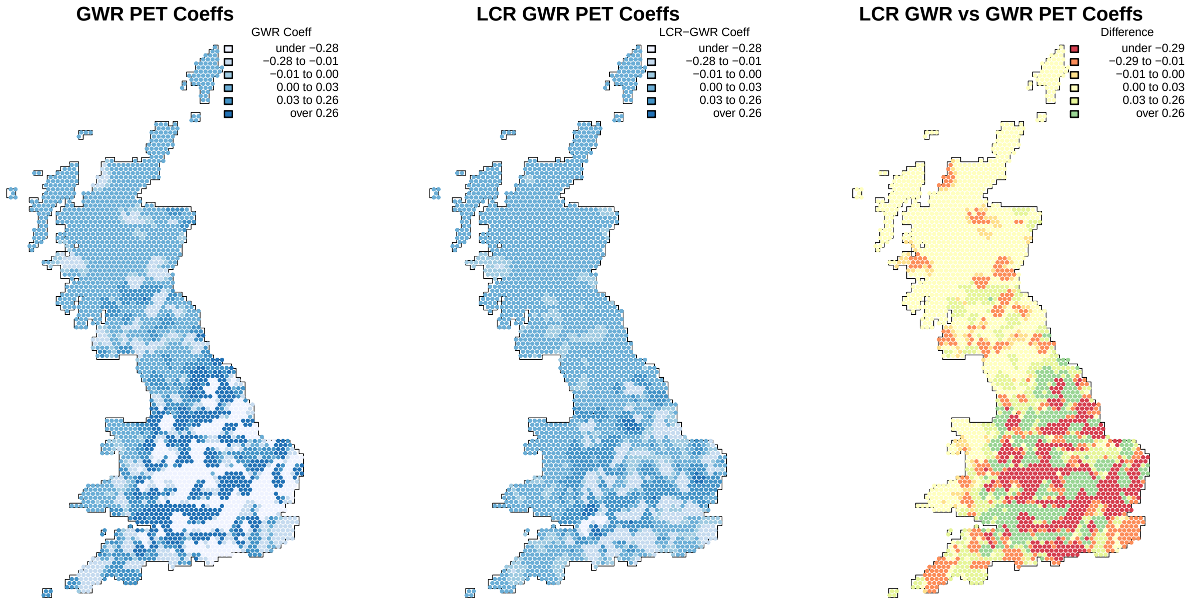

Figure 6: The coefficient estimates of the degree to which mean annual PET predicts *Acer campestre* arising from the original GWR, a locally compensated ridge GWR and a map of GWR minus LCR-GWR coefficients.

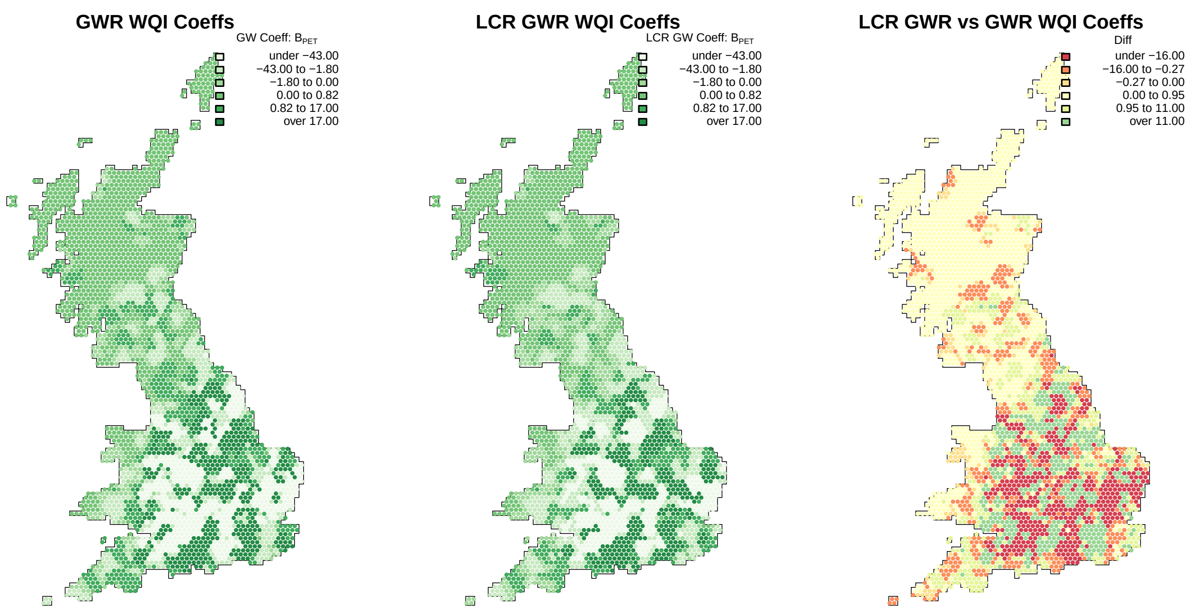

Figure 7: The coefficient estimates of the degree to which Wilderness Quality Index predicts *Acer campestre* arising from the original GWR, a locally compensated ridge GWR and a map of GWR minus LCR-GWR coefficients.

Additionally, the paper highlights the importance of considering and testing for local collinearity especially in spatial dependence models such as GWR, even where none is found to exist globally. In this analysis very strong evidence for local collinearity was found when the data were tested using local collinearity diagnostics. In most applications of GWR this critical step is missed. Where local collinearity is found a locally-compensated ridge GWR can be applied (Brunsdon et al. 2012). This only fits local ridge regressions at locations where the local CN is above a user-specified threshold. Gollini et al (2015) discuss alternative approaches for handling collinearity, but locally-compensated GWR models have the potential provide more accurate local coefficient estimates in the presence of collinearity than that found with a standard GWR model **REF needed?**.

The Geographically Weighted paradigm offers an attractive and coherent framework for many areas of applied geographical analysis. Geographically Weighted (GW) approaches develop local statistical models in order to test for spatial non-stationarity and are in contrast to standard, a-spatial, statistical approaches that assume the relationships between factors to be the same everywhere. They reflect what is commonly referred to as Tobler's 1st Law of Geography (Everything is related to everything else, but near things are more related to each other) and an understanding of the world when it is viewed through 'Geography Goggles'. These promote a vision in which the wearer is interested in how and where things vary, does not expect (statistical) relationships to be same everywhere, does not consider the world to be not normally distributed especially in space, but rather expects processes, relationships, processes, trends etc to vary spatially and to find clusters, hotspots, coldspots, etc.

These ideas are not new: quantitative geography in 1980s identified the need to move away from the whole map statistics, particularly Stan Openshaw's group at Newcastle and Julian Besag's at Durham but also Luc Anselin at Arizona. But it is important to re-state them now for a number of reasons. First, all data are spatial now (well perhaps not quite all!) but with advent of ubiquitous GPS, most records, datasets and data points have location attached to them. Second, location is not just another variable precisely because of the spatial non-stationarity observed in many processes, with the result that many phenomena are normally or randomly distributed, as predicated by classic statistical models. Third, the need to think spatially and to and to consider the spatial dimensions in a different way is given further salience by the increased access to and use of very powerful GIS software. This is increasingly resulting in instances of poor and inappropriate use of very powerful tools, but that is another story (see Comber et al., 2015). Finally, we simply observe that geography googles are not usually worn by researchers working in many areas of applied geography, especially conservation, environmental science and remote sensing, where the whole map statistic persists.

More. . . ?

6. Conclusions

TBC

Acknowledgements - TBC

Mark O'Connell Spatial Ecology and Conservation conference

References - TBC

Akaike H (1973). Information Theory and an Extension of the Maximum Likelihood Princi-ple. In BN Petrov, F Csaki (eds.), 2nd Symposium on Information Theory, pp. 267–281. Akademiai Kiado, Budapest.

Belsley DA, Kuh E, Welsch RE (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons.

Bowman A (1984). An Alternative Method of Cross-Validation for the Smoothing of Density Estimates. Biometrika, 71, 353–360.

Brunsdon C, Charlton M, Harris P (2012). Living with Collinearity in Local Regression Models. In Proceedings of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Brasil.

Brunsdon C, Fotheringham AS, Charlton M (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geographical Analysis, 28, 281–289.

Comber AJ, Brunsdon C, Green E. (2008). Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. Landscape and Urban Planning, 86: 103–114.

Comber A, Dickie J, Jarvis C, Phillips M and Tansey K, (2015). Locating bioenergy facilities using a modified GIS-based location-allocation-algorithm: considering the spatial distribution of resource supply. Applied Energy, 154: 309-316.

Coudun, C., Gégout, J.-C., Piedallu, C. and Rameau, J.-C. (2006), Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. Journal of Biogeography, 33: 1750–1763

Dormann CF et al 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27-46.

Fischer & Gosset 1935

Foody, G.M. (2004) Spatial non-stationarity and scale dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Global Ecology and Biogeography, 13, 315–320.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.

Fujita, M. (1989). Urban Economic Theory, Land Use and City Size, Cambridge: Cambridge University Press.

Gandin 1965

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geogra- phically weighted models. Journal of Statistical Software, 63 (17), 1–50.

Hijmans, Robert J., and Jane Elith. Species distribution modeling with R. (2016). http://www.idg.pl/mirrors/CRAN/web/packages/dismo/vignettes/sdm.pdf

Hoerl AE (1962). Application of Ridge Analysis to Regression Problems. Chemical Engineering Progress, 58(3), 54–59.

Hoerl AE, Kennard RW (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55–67.

Kery M., B. Gardner, and C. Monnerat, 2010. Predicting species distributions from checklist data using site-occupancy models. Journal of Biogeography. 37: 1851–1862

Kolmogorov 1941

Krige 1951

Kuiters, A. T., van Eupen, M., Carver, S., Fisher, M., Kun, Z., & Vancura, V. (2011). Wilderness register and indicator for Europe. Final Report - http://ec.europa.eu/environment/nature/natura2000/wilderness/pdf/Wilderness_register_indicator.pdf

Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. Geo-spatial Information Science, 17(2), 85-101.

Matérn 1960

Matheron 1963

Meloun, M. et al. 2002. Crucial problems in regression modelling and their solutions. Analyst 127: 433–450.

O'Brien RM (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. Quality & Quantity, 41(5), 673–690.

Phillips S.J. and J. Elith, 2011. Logistic methods for resource selection functions and presence-only species distribution models, AAAI (Association for the Advancement of Artificial Intelligence), San Francisco, USA.

Phillips, S.J., M. Dudik, J. Elith, C.H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distri- bution models: implications for background and pseudo-absence data. Ecological Applications 19: 181-197.

Tanguy, M.; Dixon, H.; Prosdocimi, I.; Morris, D. G.; Keller, V. D. J. (2015). Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2014) [CEH-GEAR]. NERC Environmental Information Data Centre. http://doi.org/10.5285/f2856ee8-da6e-4b67-bedb-590520c77b3c

Trabucco, A., and Zomer, R.J. 2009. Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www.csi.cgiar.org

VanDerWal J., L.P. Shoo, C. Graham and S.E. Williams, 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecological Modelling 220: 589- 594.

Ward G., T. Hastie, S.C. Barry, J. Elith and J.R. Leathwick, 2009. Presence- only data and the EM algorithm. Biometrics 65: 554-563.

Wheeler D (2007). Diagnostic Tools and a Remedial Method for Collinearity in Geographically Weighted Regression. Environment and Planning A, 39(10), 2464–2481.

Wheeler D (2009). Simultaneous Coefficient Penalization and Model Selection in Geographically Weighted Regression: the Geographically Weighted Lasso. Environment and Planning A, 41(3), 722–742.

Wheeler D (2013). Geographically Weighted Regression. In M Fischer, P Nijkamp (eds.), Handbook of Regional Science. Springer-Verlag.

Wheeler D, Tiefelsdorf M (2005). Multicollinearity and Correlation among Regression Co-efficients in Geographically Weighted Regression. Journal of Geographical Systems, 7(2), 161–187.

Yaglom 1955

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.