Spatial considerations in the analysis of biological conservation data: space is special

Alexis Comber[1], Steve Carver[1], Carol Ximena Garzon-Lopez[2], Paul Harris[3], Duccio Rocchini[4]

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

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

[2] Ecology and Dynamics of Human-influenced Systems Research Unit, University of Picardie Jules Verne, FR-80037 Amiens, France.

Email: c.x.garzon@gmail.com

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

Email: paul.harris@rothamsted.ac.uk

[4] Research and Innovation Centre, Dpt. Biodiversity and Molecular Ecology, Fondazione Edmund Mach, 38010 S. Michele all'Adige, Italy

Email: duccio.rocchini@fmach.it

Abstract

Geographically Weighted Regression (GWR) as first proposed by Brunsdon et al (1996) is a commonly used approach in many local geographical and biogeographical analyses. One of the of criticisms of the use of GWR is that local models may be subject to local collinearity between variables, even when none is observed globally. As yet none of the published research describing applications of GWR has addressed this issue. This paper does so. It applies GWR to the model suggested by Coudun et al (2006) to predict the presence of *Acer campestre* and to account for any spatial non-stationarity, as observed in many statistical patterns and relationships. The analysis tests for presence of local collinearity amongst predictor variables and applies a locally compensated GWR model where it is needed. The results describe the spatial variation of coefficient estimates predicting *Acer campestre* and compensates for any local collinearity amongst variables using a local ridge term. The paper discusses the need for *geography goggles* in analyses of spatial and in so doing it encourages explicitly spatial ways of thinking. Such considerations are increasingly relevant as more and more data have a spatial component in the form of location, for instance from GPS, and all data are collected some*where*. These suggest the need for more informed approaches for analysing space and location than provided by standard statistical models.

1. Introduction

This paper emphasises the importance of explicitly considering the spatial properties of ecological data and argues that location cannot be treated as just another variable because of the prevalence of spatial autocorrelation or spatial non-stationarity. To demonstrate the value of spatially explicit methods, this paper spatially extends the model suggested by Coudun et al (2006) predicting the presence of *Acer campestre* using a Geographically Weighted Regression (GWR) framework (Brunsdon et al, 1996). The extension accounts for a particular form of spatial dependence when the properties of nearby features are found to be correlated, contradicting the underlying assumptions of independence and stationarity made in many statistical analyses. The result is spatial autocorrelation or spatial non-stationarity.

Spatially explicit approaches such as GWR quantify *local* patterns in relationships and processes rather than global, 'whole map' statistics. Typically, local models are constructed from subsets of the data, for example by selecting data under a moving window or kernel, and GWR has been used in a number of ecological studies. Examples include Roll et al (2015) who examined the spatial variation tree height as a predictor of species richness for different taxa, Wang et al (2016) who used to GWR to develop local models of net primary production and Keith et al (2013) who sought to predict local species richness using a GWR analysis of environmental variables. However, local methods may be subject to collinearity, even when it is not observed globally (Brunsdon et al 2012). Collinearity affects model reliability and can result in inferential bias through unstable parameter estimates, inflated standard errors and difficulties in separating variable effects (Dormann et al 2013; Meloun et al. 2002). Despite many papers describing applications of GWR, there are no instances where the potential effects of local collinearity have been considered. As well as extending the analysis of Coudun et al (2006) using GWR, this paper explicitly tests for and compensates for locally collinearity where found, with dramatic differences between the coefficient estimates and the mapped GWR outputs without and without local compensations.

2. Background

Understanding species distribution and their spatial dependencies is a key topic in ecological research (Pearman et al., 2008; Rocchini et al., 2015) and biodiversity monitoring (Honrado et al., 2016). Reliable descriptions of distributions are fundamental for conservation and research purposes (Dormann, 2007) where a lack of spatial sensitivity hampers the ability to provide coherent biodiversity management strategies (Guisan et al., 2014, Porfirio et al., 2014). However, when modelling species distributions, a number of statistical problems must be overcome including the uncertainty related to sampling (Rocchini et al., 2011), correlation among predictors (Dormann et al., 2012), the spatial variability of predictors (Rocchini et al., 2016) and their spatio-temporal dependencies (Zuur and Ieno, 2016). Spatial non-stationarity and scale dependence have been found to have the greatest impact on observed ecological processes and patterns (Foody, 2004).

The importance of spatial considerations has a long history in a number of disciplines: Fischer (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 well recognised in socio-economic domains, but are also relevant within bio-geographical, environmental and ecological ones and are reflected in 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, despite numerous methodological advances in addressing such spatial modelling problems, whose lineage from the late 1970s can be loosely traced through spatial statistics texts from Journel and Huigbrechts (1978) to Cressie (1993), to Chilès and Delfiner (1999) and to Cressie and Wilke (2011), policy related research is still routinely informed by non-spatial modelling practices which are heroic at best, and ill-advised at worst.

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. For point-based data, geostatistical regressions are commonly used, where spatial dependencies are modelled directly via the variogram with a number of recent advances (Goovaerts 2009). 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. GWR (Brunsdon et al. 1996) was a major advance in this area. It uses a continuous distance-decay weighting scheme to accommodate the spatial structures in the data and the resultant localised regression coefficients can be mapped. Brunsdon et al. (1998) and Harris et al. (2015) provide some simple inferential mechanisms to determine whether the coefficients exhibit significant spatial variation. Some of the 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).

In summary, a key issues in ecological modelling is that dependence and some form of nonstationarity are endemic in spatial data. When spatial non-stationarity occurs the predictor and response variables are expected to change over space and the relationship between the distribution of a certain species and predictor variables may change over space. Mechanisms to handle these characteristics of spatial data in modelling is an important challenge. This paper accommodates spatial nonstationarity using the GWR model in order to demonstrate how to address process heterogeneity in relationships. The GWR model itself is formally described below, in both basic and an adapted forms 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 of *Acer campestre* presence (1980-2010) for the UK was downloaded from GBIF using the `dismo` R package. It contained 22,701 records whose spatial distribution is shown in Figure 1. The data for all years were summed over Ordnance Survey 10km grids rather than generating absence points and background data (Phillips et al. 2009; VanDerWal et al., 2009) which require a number of assumptions (Ward et al. 2009; Phillips and Elith, 2011) and may be biased (Kery et al., 2010; 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. Rainfall data were downloaded from the NERC Environmental Information Data Centre (Tanguy et al, 2015) and the average Autumn (3 month) rainfall was determined for each 1km. Mean annual potential evapotranspiration data were downloaded from the CGIAR Consortium for Spatial Information (Trabucco and Zomer, 2009) at 0.0083 degrees (approximately 1km). Finally, a wilderness quality dataset was included in the model to explore how the presence of *Acer campestre* was related to anthropogenic disturbance as anecdotally it is frequently planted as an ornamental tree and found in hedges. The European Wilderness Quality Index (WQI) (Kuiters et al 2013) provided a measure of non-anthropogenic activity. Each of these datasets were spatially aggregated over the OS 10m grid (Figure 2).

3.2 Analysis

A multi-stage approach was used to model *Acer campestre* distributions. First, a standard OLS regression to model distributions was used to identify significant predictor variables, under the assumption that relationships between predictor variables (rainfall, PET and wilderness) and species distributions are stationary (i.e. global). Second, a GWR analysis was applied to examine the spatial variation in these relationships, testing for the presence of local, non-stationarity. 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. Third, the presence of global and local collinearity amongst the predictor variables was tested for. This is a

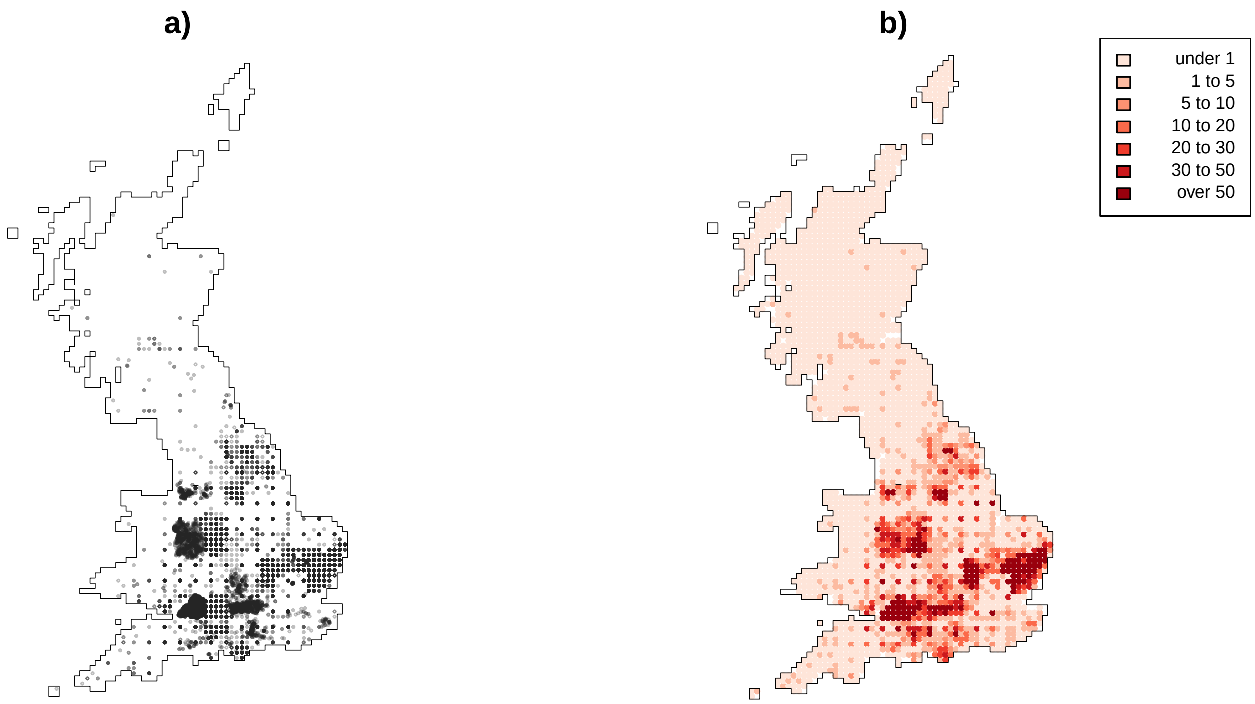

Figure 1: a) The raw data points with a transparency term to show density of points, and b) Data summed over OS 10km grid cells.

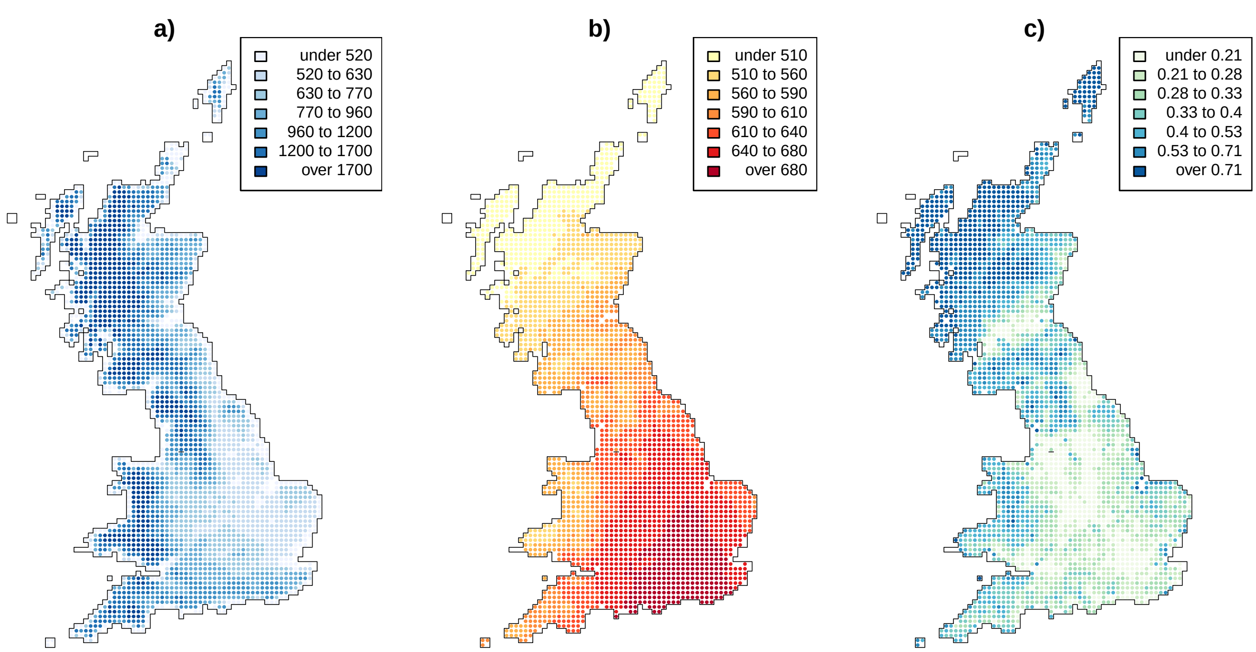

Figure 2: The data used to construct the species model, a) mean monthly Autumn Rain (mm), b) mean annual potential evapotranspiration (PET), and c) Mean Wildness Quality Index.

critical step in any regression, but for a GWR analysis it is usually overlooked. Collinearity occurs when variables exhibit linear or near linear relationships. In a GWR analyses, collinearity may occur locally, with the construction of localised regressions, even when it 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). A locally-compensated ridge GWR model (LCR-GWR) is detailed in Brunsdon et al (2012), Lu et al. (2014) and Gollini et al (2015) as are associated local collinearity diagnostics for GWR. These include local correlations amongst pairs of predictors, local Variance Inflation Factors (VIFs) for each predictor, local variance decomposition proportions (VDPs) and local condition numbers (CNs). The LCR-GWR only applies a local ridge term where it is needed: in this case when the local CN is above a pre-specified threshold of 30, which is a standard heuristic. So in summary, this study applied an OLS regression, compared this with a GWR analysis and then applied a locally-compensated ridge term if necessary. The local GWR analyses were undertaken at discrete locations in a hexagonal grid of 3028 points covering the study area.

For the geographically weighted kernels, an adaptive bi-square weighting function was applied. 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.

Optimum kernel bandwidths were found for all of the GWR models by minimising a model fit diagnostic using a leave-one-out cross-validation (CV) score (Bowman 1984; Brunsdon et al. 1996).

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 k^{th} 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 weighting results in data nearer to the kernel centre making 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

The coefficient estimates from a standard OLS regression are shown in Table 1 below. PET and Wilderness were found to be significant predictors of *Acer campestre* distribution at the 5% level. 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

Next a standard GWR analysis was applied and 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 and the significant variables, PET and WQI are mapped in Figures 3 and 4, respectively. 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.

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 (i.e. from all the data

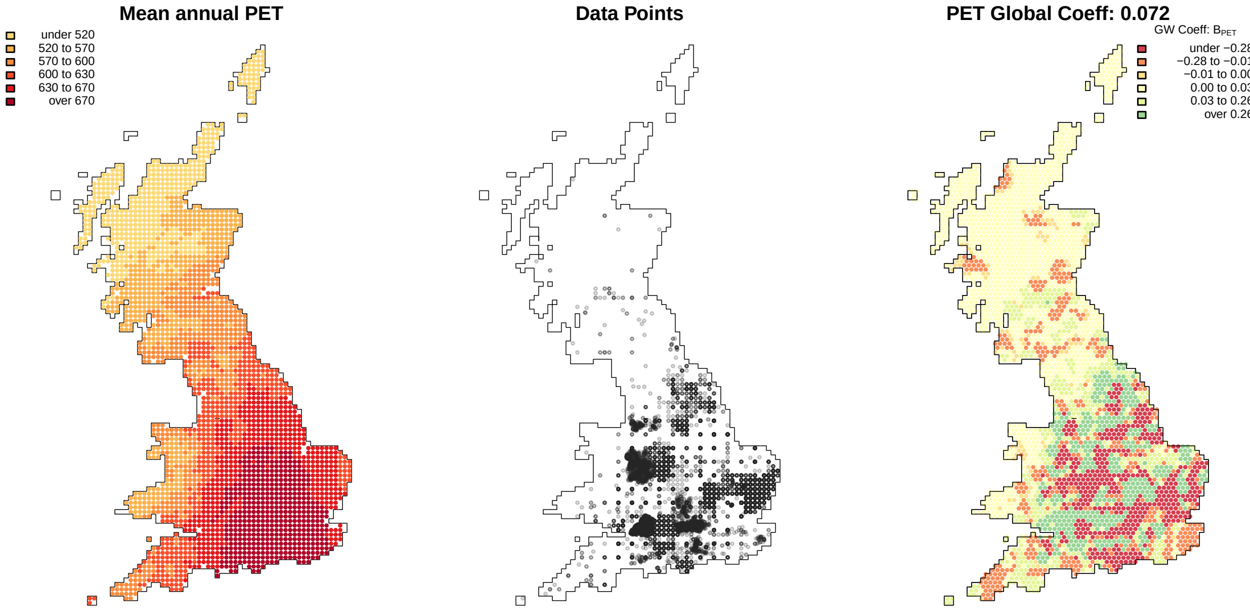

Figure 3: The spatial distribution of the mean annual PET coefficient estimates, with the context of the original PET and species data.

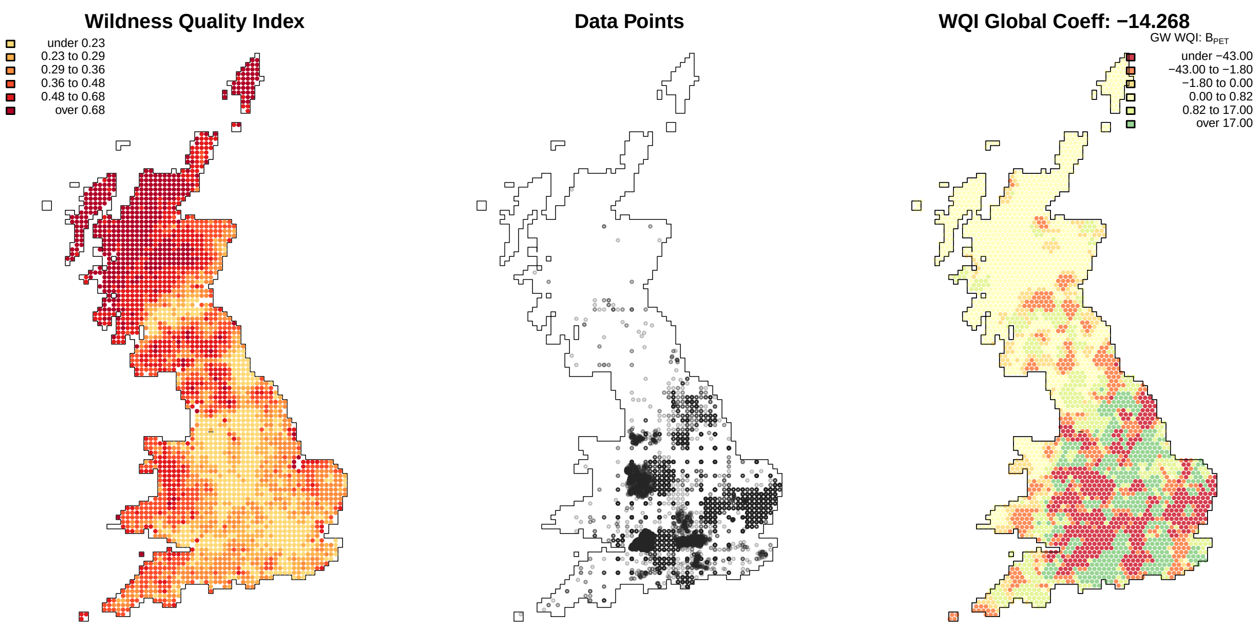

Figure 4: The spatial distribution of the Wilderness Quality Index coefficient estimates, with the context of the original WQI and species data.

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 Inflation 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). These suggest the presence of global variable collinearity. Applying GWR collinearity diagnostics to the GWR model above generates local VIFs, local VDPs and local 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 GWR model. These values suggest that the application of a LCR-GWR is warranted. The local collinearity measures are mapped in Figure 5.

Table 3. Global and local collinearity measures: Condition Numbers with Variance Inflation 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
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

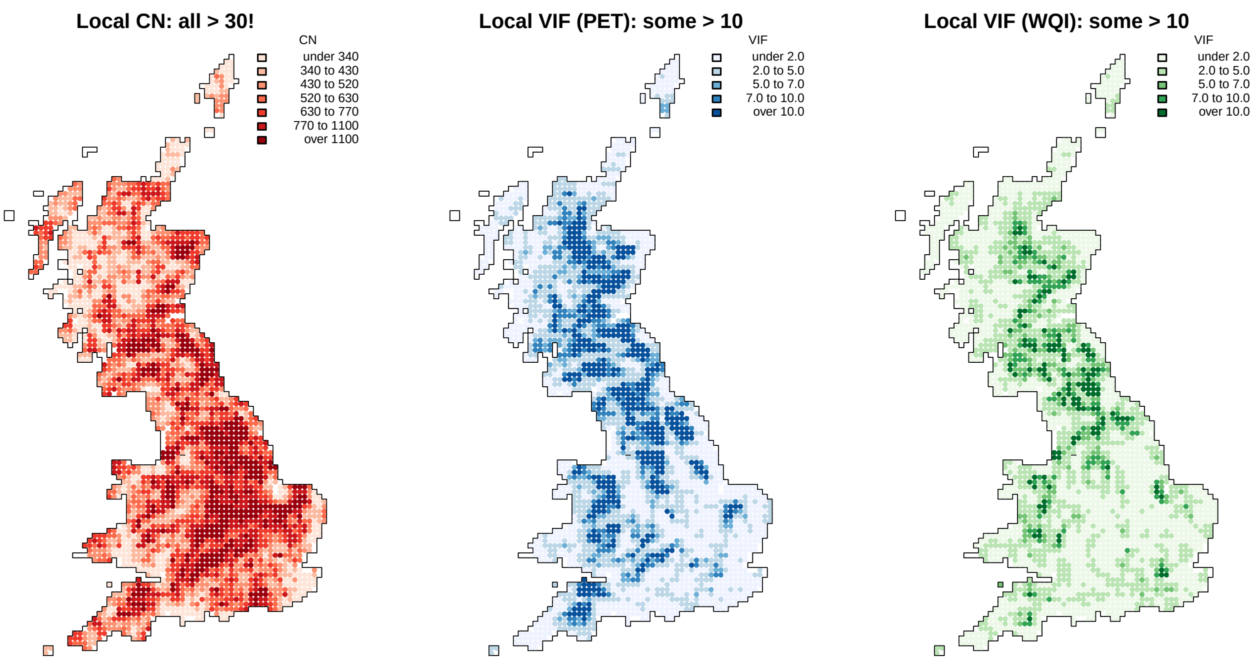

Figure 5: The spatial distribution of the local CNs and the mean annual PET VIF and the mean Wilderness Quality Index VIF.

4.3 Final GWR analysis

Having tested for and identified local collinearity, a 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 CN 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 show the spatial distribution of the original GWR coefficients, those determined using a LCR-GWR and a map of the differences between the two, for PET and for WQI. In both cases there are large and potentially important differences between the coefficient estimates from the GWR and those from the LCR-GWR.

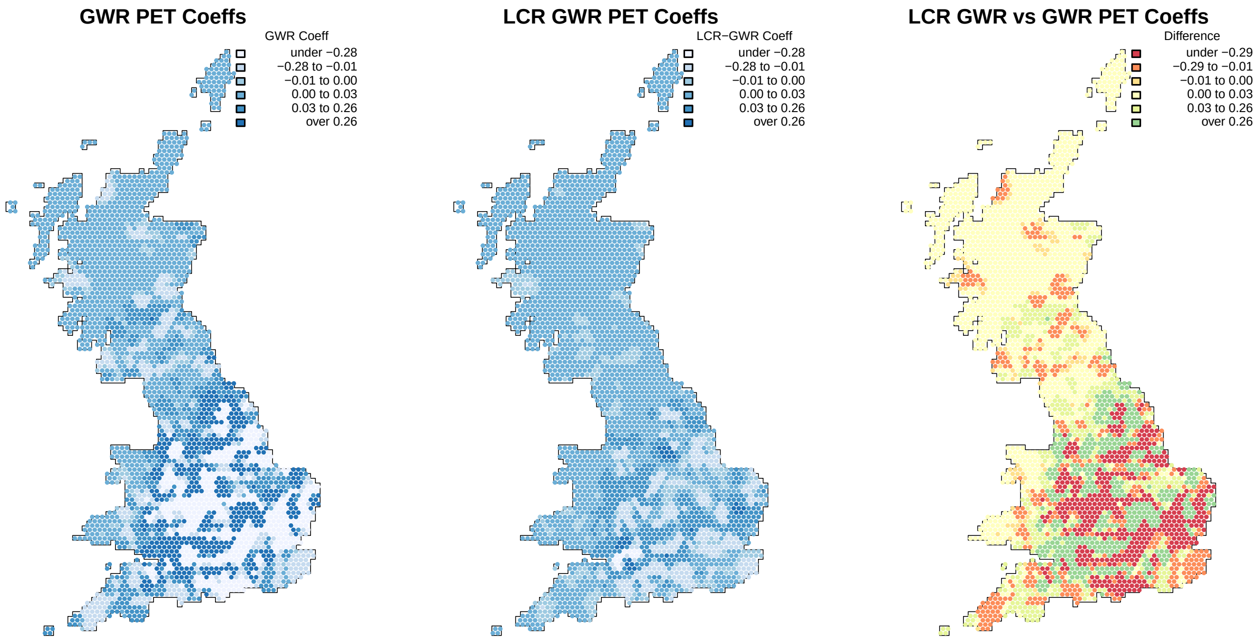

Figure 6: The coefficient estimates of the degree to which mean annual PET predicts *Acer campestre* arising from the original GWR, a locally compensated ridge GWR and a map of GWR minus LCR-GWR coefficients.

5. Discussion and Conclusions

In this paper a series of analyses were undertaken to demonstrate the application and value of explicitly spatial analyses, focusing on GWR, and the need to reconsider common assumptions in a-spatial data analyses. Local statistical models were applied to test for spatial non-stationarity, in contrast to standard, a-spatial, statistical approaches that assume the relationships between factors to be the same everywhere. Then a locally-compensated ridge GWR was used to handle any local collinearity in the GWR models, with large differences in outcome. The results highlights the importance of considering and testing for local collinearity especially in spatial non-stationarity models such as GWR, even where none is found to exist globally. This critical step has been missed in all GWR applications published to date and collinearity-aware methods were found to provide more accurate local coefficient estimates in the presence of collinearity. Gollini et al (2015) provide a review of approaches for handling local collinearity.

The Geographically Weighted (GW) paradigm offers an attractive and coherent framework for many areas of applied geographical analysis. Geographically Weighted approaches, testing for spatial non-stationarity, are in contrast to standard, a-spatial, statistical approaches that assume the relationships between factors to be the same everywhere. They reflect Tobler's 1st Law of Geography 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

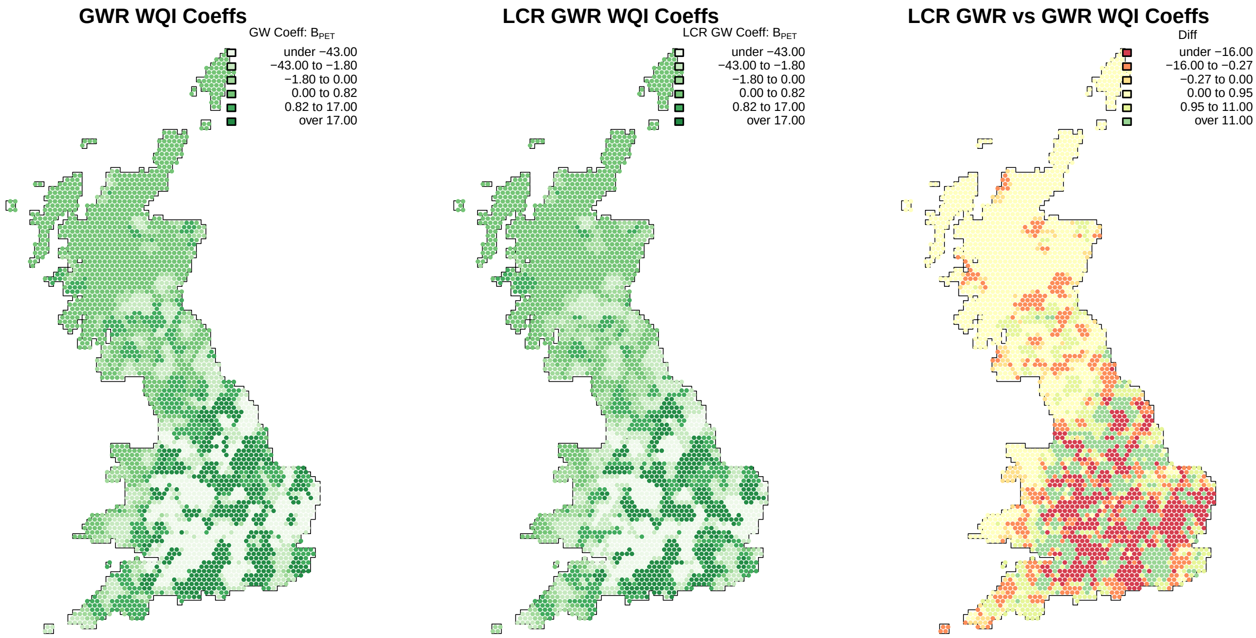

Figure 7: The coefficient estimates of the degree to which Wilderness Quality Index predicts *Acer campestre* arising from the original GWR, a locally compensated ridge GWR and a map of GWR minus LCR-GWR coefficients.

where things vary, does not expect (statistical) relationships to be same everywhere, does not consider the world to be 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, UK and Julian Besag's at Durham, UK but also Luc Anselin at Arizona, USA. Similar ideas in arose in ecology and the advent of spatial ecology. But it is important to re-state them now for a number of reasons. First, all data are spatial now and most records, datasets and data points have location attached to them with advent of ubiquitous GPS. Second, location is not just another variable precisely because of spatial heterogeneity and spatial dependence observed in many processes, with the result that many phenomena are *not* constant 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 goggles are not usually worn by researchers working in many areas of applied ecology and bio-geography, especially in conservation, environmental science and remote sensing, where the whole map statistic persists. This paper demonstrates the use of a locally compensated ridge GWR to overcome local collinearity where found. It demonstrates how carefully considered spatially explicit statistical models can be used examine the spatial variation in processes and relationships and to account for spatial autocorrelation.

Acknowledgements

The authors would like to thank Mark O'Connell and the 4th Spatial Ecology and Conservation conference for the opportunity to develop these ideas. This research was also supported by the Biotechnology and Biological Sciences Research Council (BBSRC BB/J004308/1) for the North Wyke Farm Platform.

References

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

Berry BJL, Marble DF (1968) *Spatial Analysis*, Englewood Cliffs, NJ: Prentice Hall

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

Besag, J (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B* 36: 192-225.

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.

Casetti E (1982). Drift analysis of regression parameters: An application to the investigation of fertility development relations, *Modeling & Simulation* 13: 961-966.

Chiles, J.-P. and P. Delfiner (1999). *Geostatistics - modelling spatial uncertainty.*

Chorley R, Haggett P (1967). *Models in Geography*, London: Methuen

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

Cressie, N (1993). *Statistics for Spatial Data.* New Jersey, John Wiley.

Cressie, N and Wilke CK (2011) *Statistics for Spatio-Temporal Data*, New Jersey, John Wiley.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36: 27-46.

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. and Singer, A. (2012), Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, 39: 2119–2131

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16(2), 129-138.

Fisher (1935). Statistical Tests, *Nature* 136, 474.

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.

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

Gandin, L. S. (1965). *Objective analysis of meteorological fields.*Israel Program for Scientific Translation, Jerusalem.

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

Goovaerts P (2009). Medical Geography: A promising field of application for geostatistics. *Mathematical Geosciences* 41: 243-264

Gorr WL, Olligschlaeger AM (1994). Weighted Spatial Adaptive Filtering: Monte Carlo Studies and Application to Illicit Drug Market Modeling. *Geographical Analysis*, 26: 67-87.

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16: 1424 - 1435

Harris P, Brunsdon C, Fotheringham AS (2011). Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stochastic Environmental Research and Risk Assessment* 25: 123-138

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.

Honrado, J.P., Pereira, H.M., Guisan, A. (2016), Fostering integration between biodiversity monitoring and modelling, *Journal of Applied Ecology*, 53: 1299-1304.

Journel, A. G. and C. J. Huijbregts (1978). *Mining geostatistics.* London, Academic Press.

Keith SA, Kerswell AP, and Connolly SR (2014). Global diversity of marine macroalgae: environmental conditions explain less variation in the tropics. *Global ecology and Biogeography*, 23(5), 517-529.

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, AN (1941). Interpolirovanie i ekstrapolirovanie statsionarnykh sluchainykh posledovatel' nostei (Interpolated and extrapolated stationary random sequences). Izvestya Akademiya Nauk SSSR, *Seria Matematicheskaya* 5(1).

Krige, DG (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52: 119-139.

Kuiters, A. T., van Eupen, M., Carver, S., Fisher, M., Kun, Z., & Vancura, V. (2013). *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). Spatial variation: stochastic models and their applications to problems in forest surveys and other sampling investigations. *Meddelanden fran Statens Skogforskningsinstitut*, 49: 1-144.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* 58: 1246-1266.

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.

Pearman P.B., Guisan A., Broennimann O., Randin C.F., (2008). Niche dynamics in space and time. *Trends in Ecology and Evolution*, 23: 149-158.

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.

Rocchini, D., Andreo, V., Förster, M., Garzon-Lopez, C.X., Gutierrez, A.P, Gillespie, T.W., Hauffe, H.C., He, K.S., Kleinschmit, B., Mairota, P., Marcantonio, M., Metz, M., Nagendra, H., Pareeth, S., Ponti, L., Ricotta, C., Rizzoli, A., Schaab, G., Zebisch, M., Zorer, R., Neteler, M. (2015). Potential of remote sensing to predict species invasions: A modelling perspective. *Progress in Physical Geography*, 39: 283-309.

Rocchini, D., Petras, v., Petrasova, A„ Chemin, Y., Ricotta, C., Frigeri, A., Landa, M., Marcantonio, M., Bastin, L., Metz, M., Delucchi, L., Neteler, M. (2016 in press). Spatio-ecological complexity measures in GRASS GIS. *Computers & Geosciences*.

Roll U, Geffen E, and Yom-Tov Y (2015). Linking vertebrate species richness to tree canopy height on a global scale. *Global Ecology and Biogeography*, 24(7), 814-825.

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

Tobler WR (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46: 234-240

Trabucco A, and Zomer RJ (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.

Wang Q, Ni J, and Tenhunen J (2005). Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. *Global Ecology and Biogeography*, 14(4), 379-393.

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 AM (1955). Correlation theory of processes with random stationary nth increments. *Matematicheskii Sbornik*, 37: 141-196.

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.