Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2005) **14**, 97–98

CORRESPONDENCE

Local and global approaches to spatial data analysis in ecology

Introduction

Geographic analyses in ecology may be separated into those that attempt generalizations to achieve 'global' insights, and those that attempt to explore and document local variation. Ecological studies at the broad scale usually set out to test specific hypotheses (such as the effect of energy on species richness) and focus on establishing global relationships before examining local residual variation. However, geographical pattern in model residuals (Jetz & Rahbek, 2002; Fig. 1c) can also lead to important insights. In a recent issue of *Global Ecology and Biogeography*, a study by Foody (2004) illustrates how a method for estimating local variation in model parameters, geographically weighted regression (GWR, Fotheringham *et al*., 2002), may enhance data exploration. Standard global methods, such as linear or logistic multiple regression, estimate a single parameter for each explanatory variable. In contrast, GWR allows parameter values to vary continuously in geographical space, and local parameter values are estimated by assigning higher weights to nearby observations than more distant ones. The user varies the 'bandwidth' in GWR, which determines the rate at which weights decrease with distance.

GWR

Foody uses GWR to analyse the same 1599 bird species distributions that we used to investigate the 'global' determinants of avian species richness above and beyond local variation (Jetz & Rahbek, 2001, 2002). Foody's fine-scale (small 'bandwidth') GWR analysis explained over 90% of the variation in species richness (compared to 50–60% in the global model), a value that declined towards coarser scales (larger 'bandwidths'), and most steeply so for narrow-ranged species. Single predictors such as precipitation or temperature retained high r^2 values even at coarse scales in GWR. Foody concludes that global regression models may misrepresent local conditions and yield much weaker relationships than GWR. We are concerned about the statistical and conceptual validity of this comparison, and suggest that GWR is a useful supplement but not an alternative to global modelling. In the following, we compare GWR to spatial regression models, examine Foody's conclusions, and discuss the benefits and limitations of GWR for large-scale ecology.

Non-stationarity

Observed geographical patterns and relationships in ecology, unlike physical laws that are universal, tend to be spatially variable ('non-stationary'). Even if the underlying ecological processes are universal, the realized patterns will vary with local conditions. When mapping the residuals of a traditional, non-spatial regression, they tend naturally to form 'clumps', i.e. neighbouring residuals tend to be more similar than distant ones (Legendre, 1993; Lennon, 2000; Diniz-Filho *et al*., 2003). This non-independence violates a core assumption of standard linear and logistic regression models and affects both the significance and (less appreciated) values of model parameters (Cressie, 1993; Jetz & Rahbek, 2002; Lichstein *et al*., 2002; Diniz-Filho *et al*., 2003), as well as model selection (e.g. by stepwise procedures). Failure to account for spatial autocorrelation limits the in-depth interpretation of almost all geographical analyses in ecology to date.

Spatial autocorrelation

Various techniques have been developed that incorporate the spatial covariance structure of error terms (Cliff & Ord, 1981; Cressie, 1993; Selmi & Boulinier, 2001; Lichstein *et al*., 2002). By accounting for spatial autocorrelation, these models yield unbiased statistical tests and capture much of the local variation in the response. They can thus be considered 'semi-local' (Fotheringham *et al*., 2002), even though their parameter estimates are global. To date, geographically weighted regression analyses, such as the one presented by Foody (2004), tend not to address the issue of spatial autocorrelation in model residuals, but apparently techniques to do so are under development (Fotheringham *et al*., 2002). By allowing parameters to vary locally, GWR models, particularly those with small bandwidths, are likely to capture much of the spatial pattern in the residuals. However, significant levels of autocorrelation tend to remain (Fotheringham *et al*., 2002), which may limit the validity of the regression statistics. It is important to note that while GWR accounts for 'spatial non-stationarity' in parameter estimates (the issue that some residuals are larger than others), it does not directly address autocorrelation (the issue that neighbouring residuals tend to be similar). In some cases, GWR may capture all of the spatial dependence in the residuals, but this is a rather unparsimonious method of modelling autocorrelation, as GWR models tend to have large numbers of effective parameters. In contrast, autoregressive or other models that account for autocorrelation directly (e.g. Selmi & Boulinier, 2001; Lichstein *et al*., 2002) typically require only one extra parameter compared to non-spatial models.

Model performance

Foody (2004) points out that GWR analyses explain a considerably larger proportion of the variation in species richness than conventional (global) regression, particularly at fine scales (small 'bandwidths'). This is not surprising: allowing parameters to vary locally is statistically analogous to including site factors with numerous levels (each with one degree of freedom), which precludes meaningful r^2 comparisons between GWR and global models, or between GWR models with different bandwidths. Consider the extreme case where the intercept is allowed to vary at the resolution of the data. The r^2 will be one, even without including any explanatory variables in the model. A more relevant index of model performance for GWR is the Akaike Information Criterion (AIC), which accounts for model complexity. According to AIC comparisons, GWR models did perform better than global models (Foody, 2004; Fig. 1 caption), but unfortunately, the degree to which this result may be driven by spatial autocorrelation and how it varies across variables and scale is left unexplored.

Limitations of GWR

The primary feature that distinguishes GWR, local variation in parameter values, is arguably its primary limitation for testing hypotheses. For example in Fig. 2 of Foody's paper, parameter estimates for the explanatory variables NDVI (a remotely sensed measure of the greenness of vegetation), precipitation, and temperature all range from negative to positive. Clearly, we can not make any general inferences from this analysis about how these variables affect species richness. It is possible that the effects of these variables really do vary locally, making attempts to uncover general, underlying relationships futile. It seems more likely however, that the relationships are in fact global, but appear to vary locally due to missing variables or interactions terms (e.g. the effect of temperature on species richness may switch from positive to negative depending on precipitation). Strong correlations between the local parameter estimates in Foody's Fig. 2 (NDVI and temperature parameters are positively correlated with each other, and negatively correlated with the local intercepts) are also statistically problematic and suggest that local variation in the parameters may simply reflect excessive flexibility of GWR. Rather than attempting to explain all of the variation in a response by allowing the parameters to vary locally, it seems more useful to fit global parameters, corrected for autocorrelation (Jetz & Rahbek, 2002; Lichstein *et al.*, 2002), and to appreciate that the effects of environmental factors will vary locally depending on interactions with other variables. By including interaction terms, global methods do allow for local variation in the realized effects of variables. Unlike GWR, global methods also allow one to test biological hypotheses and to make predictions that can be tested in other geographical regions (GWR can interpolate within a region, but cannot extrapolate to other regions).

Potential benefits of GWR

One benefit of GWR may be as a graphical tool for data exploration. Mapping local variation in parameter estimates may facilitate the identification of missing variables or interaction terms (e.g. variation in the effect of temperature on richness depending on local precipitation). Another potential benefit of GWR is that it provides a framework for evaluating how the strengths of relationships change with the spatial resolution of the analysis. Jetz & Rahbek (2002) found that predictability in species richness of African birds was greater for wide-ranged than for narrow-ranged species. GWR may shed additional light on this pattern. For example, Foody's Fig. 4 shows that at small bandwidth, species richness of narrow-ranged species is almost as predictable as that of wide-ranged species. As bandwidth increases, predictability declines faster for narrow compared to wide-ranged species, so that at large bandwidths, wide-ranged species are more predictable, consistent with the original analysis (Jetz & Rahbek, 2002). Just as Foody's Fig. 4 compares the scale dependence of r^2 for different response variables, one can use GWR to compare the scale dependence of r^2 for different explanatory variables with the same response (e.g. Fig. 1 in Foody, 2004). However, a potential pitfall in all of these comparisons is that the increasingly inflated GWR r^2s at small bandwidths may obscure real differences in predictability between models; i.e. at very small bandwidths, any random variable will have high r^2 in GWR. A useful way to compare r^2 across bandwidths in GWR might involve two steps: (1) Generate a large number of random variables and calculate their mean r^2 for each bandwidth. This curve then serves as a null expectation. (2) Re-scale the curves in Foody's Fig. 4 as proportional deviations from the null expectation. Variables that unlike the three mostly broad-scale predictors tested by Foody are known to act at distinctly different scales should be most fruitful for such a comparison.

Conclusions

We conclude that the local approach offered by geographically weighted regression is unsuitable for the general inference that only global models allow. Furthermore, standard GWR models do not adequately address spatial autocorrelation and must be interpreted with caution. GWR is thus not an alternative, but rather a complement to global spatial regression modelling. Its power in illustrating local performance of predictor variables and their interaction with scale makes it a useful tool for ecological analyses at the broad scale.

Walter Jetz,[1,2,*] Carsten Rahbek[3] and Jeremy W. Lichstein[1]

[1]*Ecology and Evolutionary Biology Department, Princeton University, 08544 Princeton NJ, USA;* [2]*Biology Department, University of New Mexico, 87131 Albuquerque NM, USA. E-mail: wjetz@princeton.edu;* [3]*Zoological Museum, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen O Denmark. E-mail: crahbek@zmuc.ku.dk*

**Correspondence: Walter Jetz, Ecology and Evolutionary Biology Department, Princeton University, 08544 Princeton NJ, USA. E-mail: wjetz@princeton.edu*

ACKNOWLEDGEMENTS

We thank Giles M. Foody, David. J. Rogers, Richard Field and an anonymous referee for comments on the manuscript.

REFERENCES

Cliff, A.D. & Ord, J.K. (1981) *Spatial processes: models and applications*. Pion, London.

Cressie, N.A.C. (1993) *Statistics for spatial data*. Revised edn. Wiley, New York; Chichester.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.

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*. Wiley, Chichester; Hoboken, N.J.

Jetz, W. & Rahbek, C. (2001) Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5661–5666.

Jetz, W. & Rahbek, C. (2002) Geographic range size and determinants of avian species richness. *Science*, **297**, 1548–1551.

Legendre, P. (1993) Spatial autocorrelation — trouble or new paradigm. *Ecology*, **74**, 1659–1673.

Lennon, J.J. (2000) Red shifts and red herrings in geographical ecology. *Ecography*, **23**, 101–113.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E. (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445–463.

Selmi, S. & Boulinier, T. (2001) Ecological biogeography of Southern Ocean islands: the importance of considering spatial issues. *American Naturalist*, **158**, 426–437.