Habitat Association Modelling for Farmland Birds

Report to the Economics Team of the UK National Ecosystem Assessment

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

December 2010

Author: Stephen Dugdale

Introduction

This report outlines the methodology and results of habitat association modelling for a suite of 19 farmland birds in England and Wales. All 19 species belong to the same guild as they all have a diet of seeds and invertebrates during the breeding season. Details of the species included are shown in Appendix 1.

Method

The bird distribution data was collected by volunteers for the British Trust for Ornithology (BTO) between 1988 and 1991 and published in The New Atlas of Breeding Birds in Britain and Ireland : 1988-1991 (Gibbons, Reid & Chapman, 1993). The dependent variable is guild richness; this being the sum of the number of species from the seeds and invertebrates guild present at each location.

The spatial resolution is 10km squares in England and Wales based on the British National Grid. Any square with less than 25% of agricultural land in 1988 based on the Agricultural Census data (EDINA: available at http://edina.ac.uk/agcensus/) was excluded from the analysis. A total of 1,505 squares matched these criteria. However, bird data were only available for 1,496 of these squares and therefore the habitat association models are based on this slightly reduced area. Excluded squares were either coastal squares with a significant part of the land area potentially offshore, or part of the major conurbations. Figure 1 shows details of the study area and excluded squares.

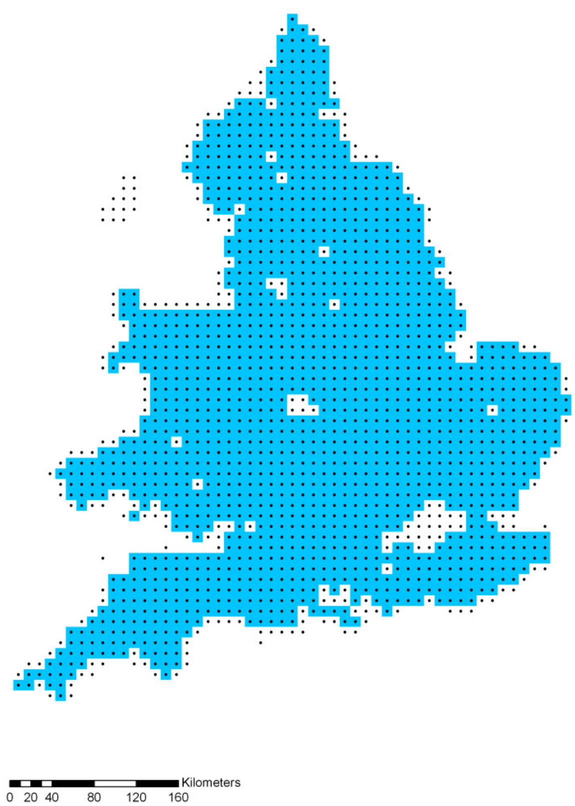

Figure 1: The distribution of 10km squares comprising the study area. Squares not highlighted in blue were excluded from the analysis due to having less than 25% agricultural land.

Fifteen habitat variables were initially selected to be utilised in the habitat association models (Table 1). In order to avoid multicollinearity among variables all pairwise correlations were examined to identify correlated pairs ($r > 0.7$) (Caprio, Ellena & Rolando, 2009; Siriwardena et al., 2000). All except one of any pair or group of variables where $r > 0.7$ was removed (Caprio, Ellena & Rolando, 2009; Siriwardena et al., 2000). Selection of which variable to retain was mostly based on a subjective estimate of interpretability (Caprio, Ellena & Rolando, 2009). However, where highly correlated variables were equally interpretable, other criteria used were to give preference to variables that were more stable over time (e.g. altitude over Ag. Census data), or variables that were more widespread (e.g. NEA_Cereal over rape).

Table 1: Initial explanatory variables. Shaded rows were removed due to multicollinearity.

Explanatory Variable	Explanation	Data Source
nea_cereal	Wheat + Summer Barley + Winter Barley + Oats	1988 Agricultural Census
nea_root	Potatoes + Sugar Beet	1988 Agricultural Census
rape_88	Rape	1988 Agricultural Census
tempgras88	Temporary Grass	1988 Agricultural Census
permgras88	Permanent Grass	1988 Agricultural Census
rough_88	Rough Grazing	1988 Agricultural Census
tot_cat_88	Total Cattle (000's)	1988 Agricultural Census
tot_shp_88	Total Sheep (000's)	1988 Agricultural Census
conif_wood	Coniferous Woodland	Land Cover Map (1990)
decid_wood	Deciduous Woodland	Land Cover Map (1990)
avg_temp [1]	Average Temperature	??? from Carlo
Rainfall [2]	Rainfall	??? from Carlo
urban_pc	Percentage of urban land	??? from Carlo
alt_range	Range between maximum and minimum altitude (m)	Calculated *
alt_mean	Mean altitude (m)	Calculated *

[1] 30 year (1961 – 1991) average temperature in the growing season (April – September)
[2] Total rainfall calculated as the 30 year average from 1961 – 1991
* Source data were OS Panorama DEM

All remaining variables were placed into an initial Ordinary Least Squares (OLS) model and variables sequentially removed in order of the lowest significance. In this study modern model selection techniques that are increasingly advocated for use in ecological modelling were used (Johnson & Omland, 2004). At each step, model performance was assessed by Akaike's information criterion (AIC) as AIC aids in identifying the most parsimonious model amongst a set and avoids having to rely on significance levels which are by definition arbitrary (Rushton, Ormerod & Kerby, 2004). As a general rule, improvements in the AIC that are less than 3 in value could easily arise as a result of sampling error, whereas values of greater than 3 are more likely to be due to genuine differences in the models (Fotheringham, Brunsdon & Charlton, 2002). Therefore, the "best" OLS model was that with the fewest parameters amongst those that had an AIC value within 3 of the model with the overall lowest AIC score. Finally, the model was tested for spatial autocorrelation by calculating Morans I statistic (Moran, 1950) using the spatial statistics tool for that purpose in ARCMAP 9.3.

A Geographically Weighted Regression (GWR) model was also constructed using a bandwidth of 33km. The bandwidth parameter was based on the mean dispersal distance for all species in the guild for which data were available. Appendix 2 of Wernham et al. (2002) gives separate details of breeding and natal dispersal distances and their standard deviations by species. To take account of the variability in the data an "upper breeding dispersal distance" was calculated as breeding dispersal distance + 1.96 standard deviations for each species in the guild. The mean value of the "upper breeding dispersal distance" for the guild was then calculated. Exactly the same calculation was then performed using the natal dispersal distance data. The bandwidth for a guild was set to the greater of the mean upper breeding dispersal distance or the mean upper natal dispersal distance.

Results

The dependent variable (guild richness) had a theoretical range of 0 to 19 species. In practice the range was 6 to 18 species which was distributed spatially as shown in Figure 2.

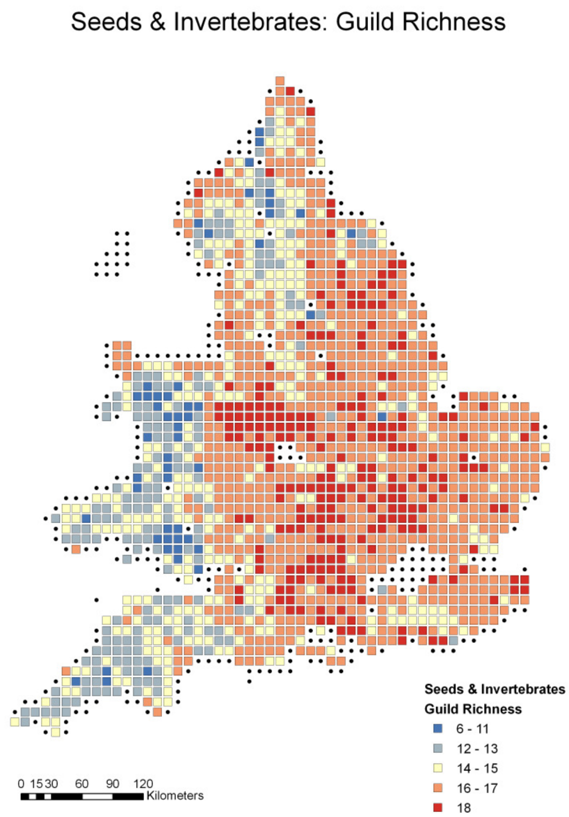

Figure 2: Spatial distribution of guild richness for the Seeds and Invertebrates guild in 1991.

The results of OLS regression models were unreliable (as explained below) and therefore additional analysis was conducted using Geographically Weighted Regression (GWR).

Ordinary Least Squares (OLS) Regression Results

Regression coefficients of statistically significant variables for the most parsimonious model are shown in Table 2. However, analysis of the diagnostics indicated that the results from this model are not reliable for two reasons. Firstly, there is evidence of statistically significant levels of heteroscedasticity and/or non-stationarity of the data (Koenker (BP) Statistic = 58.18, $p < 0.001$). The model was significantly underestimating guild richness in South-West England and South

Wales and overestimating guild richness in the West Midlands (Figure 3) confirming that non-stationarity of the data was an issue. Spatial autocorrelation was also present in the residuals ($Z = 36.66$, $p < 0.001$) indicating a significant level of clustering. The residuals were also not normally distributed (Jarque-Bera Statistic = 174.47, $p < 0.001$).

Table 2: Regression coefficients and significance levels for variables in the best OLS model.

Explanatory Variable	Coefficient	Sig.
Intercept	14.736	< 0.001
nea_cereal	0.049	< 0.001
tempgras88	0.033	< 0.001
conif_wood	-0.066	< 0.001
urban_pc	0.029	< 0.001
alt_mean	-0.004	< 0.001

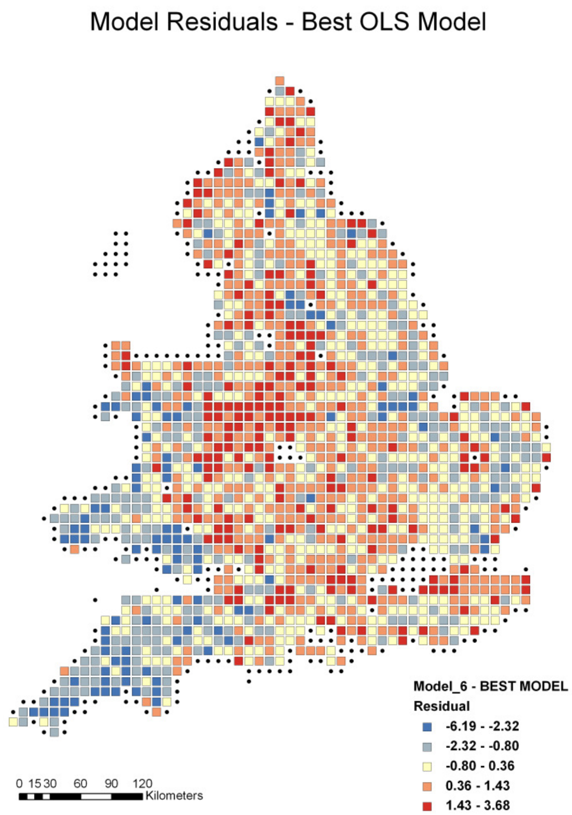

Figure 3: Spatial distribution of the residuals from the best OLS model.

Geographically Weighted Regression (GWR) Results

Geographically Weighted Regression (GWR) provides an alternative, and perhaps more logical, solution to the problem of spatially auto-correlated error terms in spatial modelling compared with the various forms of spatial regression modelling (Fotheringham, Brunsdon & Charlton, 2002). The GWR methodology was therefore applied to the same dependent and independent variables as the "best" OLS model. Instead of a single regression coefficient for each variable (as produced in OLS models), GWR provides a regression coefficient for each variable at every data point. The spatial variation in the regression coefficients can then be mapped and displayed visually. Similarly, model performance (r^2) is also calculated at each data point.

Residuals from the GWR model were not spatially autocorrelated ($Z = -0.58$, $p = 0.56$). The spatial distribution of the residuals is shown in Figure 4.

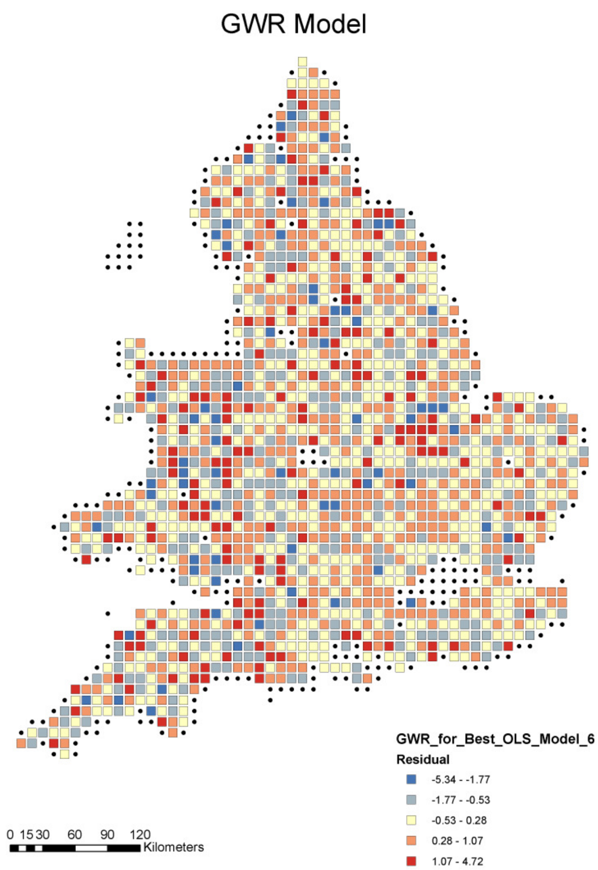

Figure 4: Spatial distribution of the residuals for the GWR model (classification by Jenks Natural Breaks).

Model performance differed considerably across the study area (Table 3 and Figure 5) with the best model performance being in the north of England and north Wales.

Table 3: Model performance (r^2) and the proportion of the study area for each 10% band

	No.	**Prop'n**
Prop'n Squares $r^2 < 0.10$	245	0.16
Prop'n Squares $r^2 = 0.10 – 0.20$	293	0.20
Prop'n Squares $r^2 = 0.20 – 0.30$	249	0.17
Prop'n Squares $r^2 = 0.30 – 0.40$	197	0.13
Prop'n Squares $r^2 = 0.40 – 0.50$	150	0.10
Prop'n Squares $r^2 = 0.50 – 0.60$	185	0.12
Prop'n Squares $r^2 = 0.60 – 0.70$	113	0.08
Prop'n Squares $r^2 = 0.70 – 0.80$	64	0.04
Prop'n Squares $r^2 = 0.80 – 0.90$	0	0.00
Prop'n Squares $r^2 = 0.90 – 1.00$	0	0.00

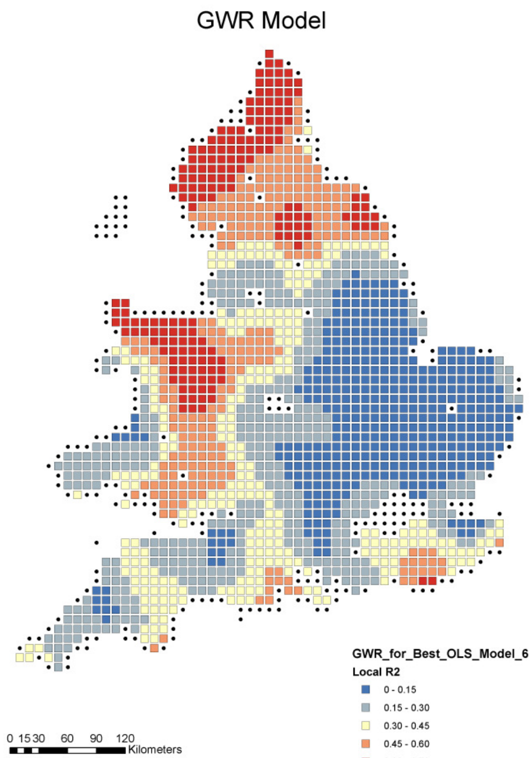

Figure 5: The spatial distribution of model performance (r^2).

Each predictor variable had both positive and negative effects on guild richness in different parts of the study area. The percentage split between positive and negative effects is shown in Table 4 and the range of the regression coefficients is shown in Table 5. The spatial distribution of the variation in the predicted effect on guild richness is shown for the intercept (Figure 6a) and for each of the predictor variables separately (Figure 6b to 6f).

Table 4: The percentage of the overall study area (1,496 squares) displaying positive and negative effects for each predictor variable.

Variable	Negative Effect	Positive Effect
nea_cereal	22.79	77.21
tempgras88	42.98	57.02
conif_wood	65.71	34.29
urban_pc	39.17	60.83
alt_mean	73.80	26.20

Table 5: The range of regression coefficients for each predictor variable

Variable	Range of regression coefficients
nea_cereal	-0.16 - 0.33
tempgras88	-0.34 - 0.38
conif_wood	-0.38 - 1.99
urban_pc	-0.22 - 0.12
alt_mean	-0.03 - 0.01

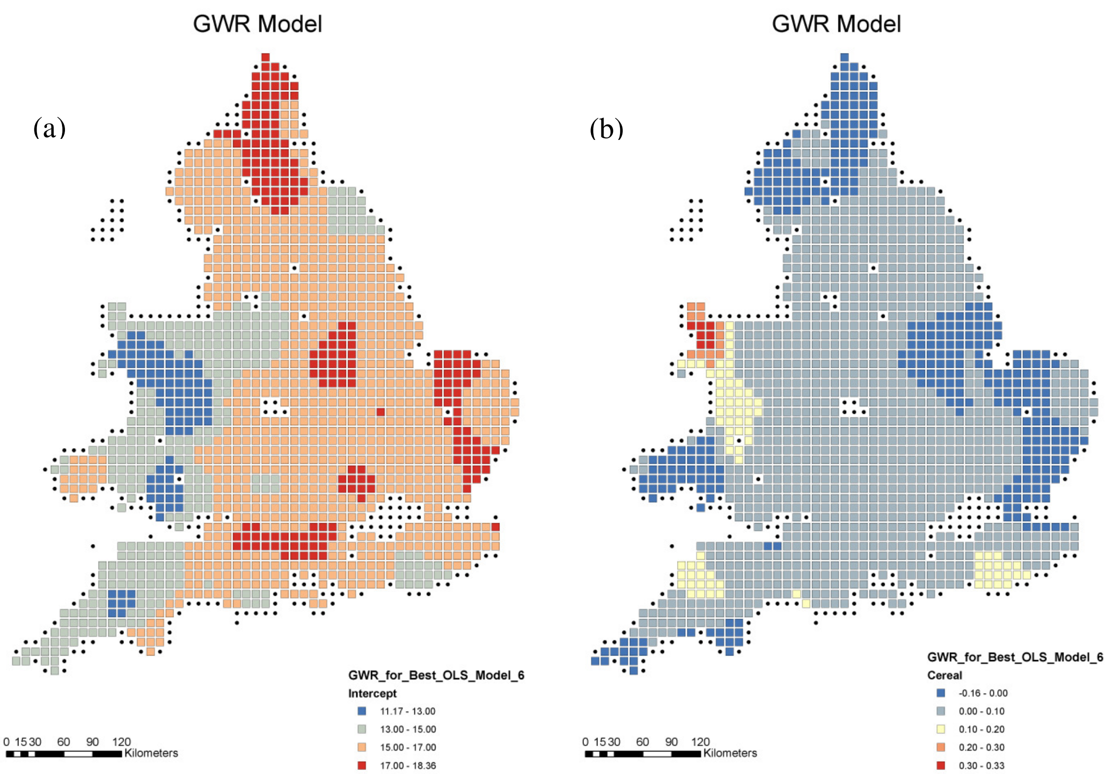

Figure 6: Spatial variation of the value of the intercept (a) and the effect on guild richness for the Seeds and Invertebrates guild for each of the predictor variables (b) cereal, (c) temporary grass, (d) coniferous woodland, (e) urban and (f) mean altitude (continued on next page).

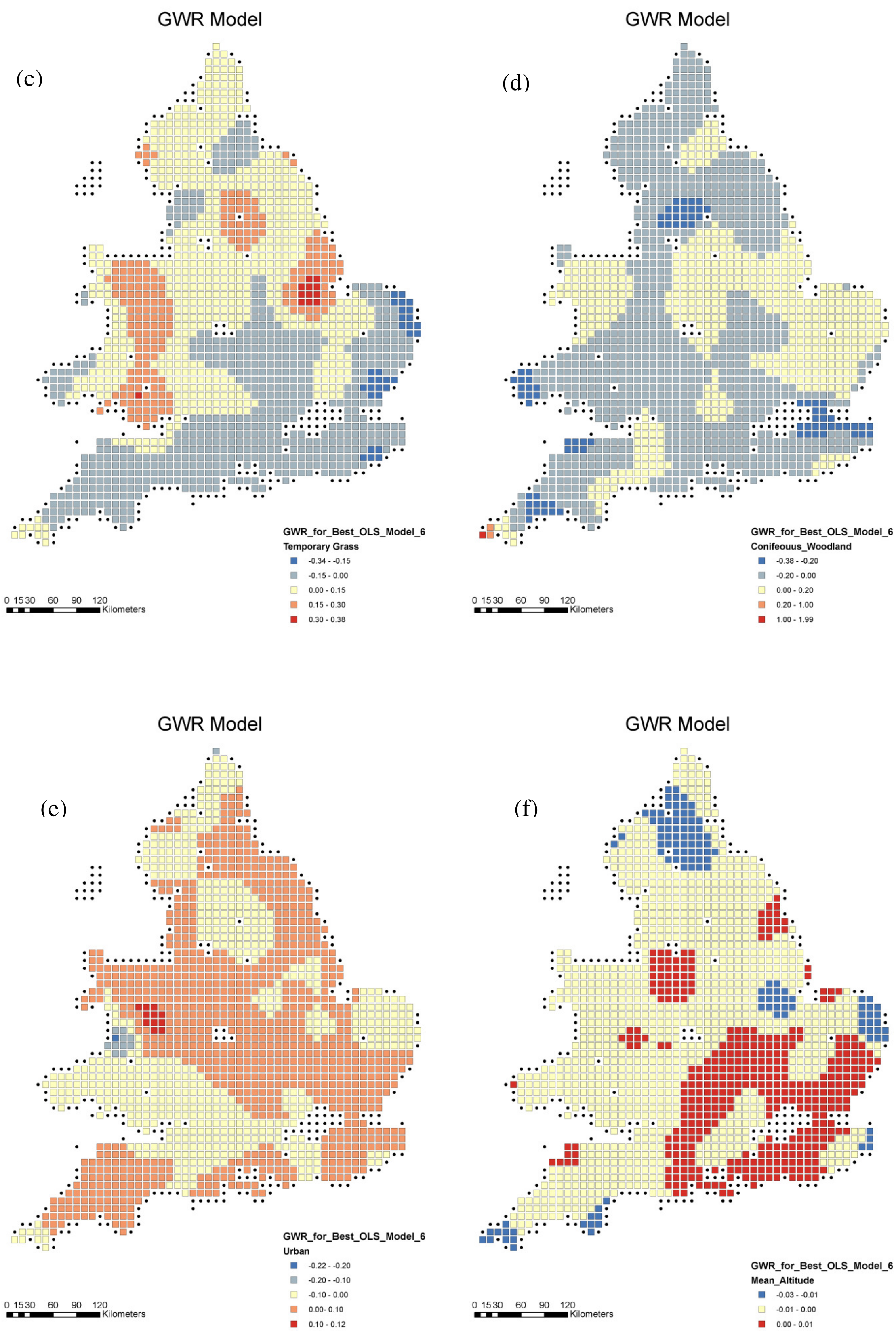

Figure 6 (continued). See caption on previous page.

Discussion

The results from the GWR model are robust as the methodology successfully removed spatial autocorrelation in the data. Model performance was good for an ecological model (r >= 0.30) in 47% of the study area. Model performance tended to be better in upland areas (Figure 4) suggesting that one or more important predictor variables may have been excluded from the model. However, in the absence of additional suitable data to revise the models, the results will be discussed as they are the best achievable models at the current time.

All predictor variables had both positive and negative effects on guild richness depending on location. Cereals and temporary grass had positive effects on guild richness in more than 50% of the study areas which is not unexpected. However, as the amount of urban land increased, an increase in guild richness was also predicted in more than 50% of the study area. A possible explanation for this is that several of the species are also found in gardens or parkland in urban areas. Both coniferous woodland and mean altitude had negative effects on guild richness in more than 50% of the study area (Table 4).

The regression coefficient for mean altitude was very small as the unit of measurement was per metre. If this was rescaled to per 100 metres, then the effect of altitude was of a similar order to the other predictor variables. Coniferous woodland had large positive regression coefficients of up to 1.99. However, only three 10km squares had values greater than 0.2 and all of these were at or adjacent to Lands End (Figure 6d) and therefore it is suggested that these are outliers which are considerably different from predicted values for the rest of the study area.

Conclusion

England and Wales is a large study area with a diversity of landscapes and agricultural regimes. Use of global regression models such as OLS tend to average out variations in regression coefficients in order to produce a model of best fit. Consequently, there is the risk that the global regression coefficients are not representative of any location within the study area. The OLS diagnostics suggest that this has occurred in the OLS model produced within this study, and therefore the results cannot be regarded as reliable.

The GWR model confirms the existence of regional variation in regression coefficients and was successful in removing spatial autocorrelation. Therefore the GWR results can be regarded as much more robust. However, model performance varied considerably and showed a distinct upland/lowland split between good and poor model performance. It is therefore suggested that the GWR results are treated with some caution and that a more detailed initial set of predictor variables could produce more meaningful results.

However, for the purpose of the National Ecosystem Assessment where the purpose is to estimate general trends in the change in guild richness based on scenarios of changing land use, the GWR results are likely to be fit for purpose.

References

Caprio, E., Ellena, I. & Rolando, A. (2009) Assessing habitat/landscape predictors of bird diversity in managed deciduous forests: a seasonal and guild-based approach. *Biodiversity and Conservation*, **18**(5), 1287-303.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002) *Geographically Weighted Regression - the analysis of spatially varying relationships* John Wiley & Sons, Chichester.

Gibbons, D.W., Reid, J.B. & Chapman, R.A. (1993) *The New atlas of breeding birds in Britain and Ireland : 1988-1991* T. & A.D. Poyser, London.
Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution*, **19**(2), 101-08.
Moran, P.A.P. (1950) Notes On Continuous Stochastic Phenomena. *Biometrika*, **37**(1-2), 17-23.
Rushton, S.P., Ormerod, S.J. & Kerby, G. (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**(2), 193-200.
Siriwardena, G.M., Crick, H.Q.P., Baillie, S.R. & Wilson, J.D. (2000) Agricultural land-use and the spatial distribution of granivorous lowland farmland birds. *Ecography*, **23**(6), 702-19.
Wernham, C.V., Toms, M.P., Marchant, J.H., Clark, J.A., Siriwardena, G.M. & Baillie, S.R., eds. (2002) *The Migration Atlas: movements of the birds of Britain and Ireland* T. & A.D. Poyser, London.

Appendix 1: Farmland Bird Species in the Seeds and Invertebrates Guild

Common Name	Latin Name	BTO Code
Bullfinch	*Pyrrhula pyrrhula*	BF
Carrion Crow	*Corvus corone corone*	C
Chaffinch	*Fringilla coelebs*	CH
Cirl Bunting	*Emberiza cirlus*	CL
Corn Bunting	*Miliaria calandra*	CB
Goldfinch	*Carduelis carduelis*	GO
Greenfinch	*Carduelis chloris*	GR
Grey Partridge	*Perdix perdix*	P
Jackdaw	*Corvus monedula*	JD
Linnet	*Carduelis cannabina*	LI
Magpie	*Pica pica*	MG
Pheasant	*Phasianus colchicus*	PH
Quail	*Coturnix coturnix*	Q
Red-legged Partridge	*Alectoris rufa*	RL
Reed Bunting	*Emberiza schoeniclus*	RB
Rook	*Corvus frugilegus*	RO
Skylark	*Alauda arvensis*	S
Tree Sparrow	*Passer montanus*	TS
Yellowhammer	*Emberiza citronella*	Y