

REVIEW AND SYNTHESIS

Regression analysis of spatial data

Colin M. Beale,^{1††} Jack J. Lennon,¹ Jon M. Yearsley,^{2,3} Mark J. Brewer⁴ and David A. Elston⁴

¹The Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH, UK

²Département d'Écologie et Évolution, Université de Lausanne, CH-1015 Lausanne, Switzerland

³School of Biology & Environmental Science, UCD Science Centre, Belfield, Dublin 4, Ireland

⁴Biomathematics and Statistics Scotland, Craigiebuckler, Aberdeen, AB15 8QH, UK

*Correspondence: E-mail: cb751@york.ac.uk

†Present address: Department of Biology (Area 18), PO Box 373, University of York, YO10 5YW

Abstract

Many of the most interesting questions ecologists ask lead to analyses of spatial data. Yet, perhaps confused by the large number of statistical models and fitting methods available, many ecologists seem to believe this is best left to specialists. Here, we describe the issues that need consideration when analysing spatial data and illustrate these using simulation studies. Our comparative analysis involves using methods including generalized least squares, spatial filters, wavelet revised models, conditional autoregressive models and generalized additive mixed models to estimate regression coefficients from synthetic but realistic data sets, including some which violate standard regression assumptions. We assess the performance of each method using two measures and using statistical error rates for model selection. Methods that performed well included generalized least squares family of models and a Bayesian implementation of the conditional auto-regressive model. Ordinary least squares also performed adequately in the absence of model selection, but had poorly controlled Type I error rates and so did not show the improvements in performance under model selection when using the above methods. Removing large-scale spatial trends in the response led to poor performance. These are empirical results; hence extrapolation of these findings to other situations should be performed cautiously. Nevertheless, our simulation-based approach provides much stronger evidence for comparative analysis than assessments based on single or small numbers of data sets, and should be considered a necessary foundation for statements of this type in future.

Keywords

Conditional autoregressive, generalized least squares, macroecology, ordinary least squares, simultaneous autoregressive, spatial analysis, spatial autocorrelation, spatial eigenvector analysis.

Ecology Letters (2010) 13: 246–264

INTRODUCTION

With the growing availability of remote sensing, global positioning services and geographical information systems many recent ecological questions are intrinsically spatial: for example, what do spatial patterns of disease incidence tell us about sources and vectors (Woodroffe *et al.* 2006; Carter *et al.* 2007; Jones *et al.* 2008)? How does the spatial scale of human activity impact biodiversity (Nogues-Bravo *et al.* 2008) or biological interactions (McMahon & Diez 2007)? How does the spatial structure of species' distribution patterns affect ecosystem services (Wiegand *et al.* 2007; Vandermeer *et al.* 2008)? Can spatially explicit conservation

plans be developed (Grand *et al.* 2007; Pressey *et al.* 2007; Kremen *et al.* 2008)? Are biodiversity patterns driven by climate (Gaston 2000)? While many ecologists recognize that there are special statistical issues that need consideration, they often believe that spatial analysis is best left to specialists. This is not necessarily true and may reflect a lack of baseline knowledge about the relative performance of the methods available.

A plethora of new spatial models are now available to ecologists, but while discrepancies between the models and their fitting methods have been noted (e.g. Dormann 2007), it is essentially unknown how well these different methods perform relative to each other, and consequently what are

their strengths and weaknesses. For example, application of several methods to a single data set can lead to regression coefficients that actually differ in sign as well as in magnitude and significance level for a given explanatory variable (Beale *et al.* 2007; Diniz-Filho *et al.* 2007; Dormann 2007; Hawkins *et al.* 2007; Kühn 2007). Indeed, recent applications of a range of spatial regression methods to an extensive survey of real datasets concluded that the difference in regression coefficients between spatial (allowing for autocorrelation) and non-spatial (i.e. ordinary least squares) regression analysis is essentially unpredictable (Bini *et al.* 2009). It is perhaps this confusion that explains why a recent review of the ecological literature found that 80% of studies analysing spatial data did not use spatially explicit statistical models at all, despite the potential for introducing erroneous results into the ecological literature if important features of the data are not properly accounted for in the analysis (Dormann 2007). It follows that the important synthesis required by ecologists is the identification of which methods consistently perform better than others when applied to real data sets.

Unfortunately, in real-world situations it is impossible to know the true relationships between covariates and dependent variables (Miao *et al.* 2009), so performance of different modelling techniques can never be convincingly assessed using real data sets. In other words, without controlling the relationships between and properties of the response variable, y , and associated explanatory, x , variables, the relative ability of a suite of statistical tools to estimate these relationships is impossible to quantify: one can never know if the results are a true reflection of the input data or an artefact of the analytical method. Here, we measure how well each method performs in terms of bias (systematic deviation from the true value) and precision (variation around the true value) of parameter estimates by using a series of scenarios in which the relationships are linear, the explanatory variables exhibit spatial patterns and the errors about the true relationships exhibit spatial auto-correlation. These scenarios describe a range of realistic complexity that may (and is certainly often assumed to) underlie ecological data sets, allowing the performance of methods to be assessed when model assumptions are violated as well as when model assumptions are met. By using multiple simulations from each scenario, we can compare the true value with the distribution of parameter estimates: such an approach has been standard in statistical literature since the start of the 20th century (Morgan 1984) and has also been used in similar ecological contexts (e.g. Beale *et al.* 2007; Dormann *et al.* 2007; Carl & Kuhn 2008; Kissling & Carl 2008; Beguería & Pueyo 2009). Previously, however, such studies have been limited both in the lack of complexity of the simulated datasets and by the limited range of tested methods (Kissling & Carl 2008; Beguería & Pueyo 2009) or

data sets (e.g. Dormann *et al.* 2007) or both (Beale *et al.* 2007; Carl & Kuhn 2008). Here, we describe simulations and analyses that overcome these previous weaknesses and so significantly advance our understanding of methods to use for spatial analysis.

Highly detailed reference books have been written on analytical methods for the many different types of spatial data sets (Haining 1990, 2003; Cressie 1993; Fortin & Dale 2005) and we do not attempt an extensive review. Instead, we provide a comparative overview and an evidence base to assist with model and method selection. We limit ourselves to linear regression, with spatially correlated Gaussian errors, the most common spatial analysis that ecologists are likely to encounter and a relatively straightforward extension of the statistical model familiar to most. The approach we take and many of the principles we cover, however, are directly relevant to other spatial analysis techniques.

Why is space special?

Statistical issues in spatial analysis of a response variable focus on the almost ubiquitous phenomenon that two measurements taken from geographically close locations are often more similar than measurements from more widely separated locations (Hurlbert 1984; Koenig & Knops 1998; Koenig 1999). Ecological causes of this spatial autocorrelation may be both extrinsic and intrinsic and have been extensively discussed (Legendre 1993; Koenig 1999; Lennon 2000; Lichstein *et al.* 2002). For example, intrinsic factors (aggregation and dispersal) result in autocorrelation in species' distributions even in theoretical neutral models with no external environmental drivers of species distribution patterns. Similarly, autocorrelated extrinsic factors such as soil type and climate conditions that influence the response variable necessarily induce spatial autocorrelation in the response variable (known as the Moran effect in population ecology). Whilst these processes usually lead to positive autocorrelations in ecological data, they may also generate negative autocorrelations, when near observations are more dissimilar than more distant ones. Negative autocorrelation can also occur when the spatial scale of a regular sampling design is around half the scale of the ecological process of interest. As ecological examples of negative autocorrelation are rare and the statistical issues similar to those of positive autocorrelation (Velando & Freire 2001; Karagatzides *et al.* 2003), all the scenarios we consider have positive autocorrelation.

The potential for autocorrelation to vary independently in both strength and scale is often overlooked (Cheal *et al.* 2007; Saether *et al.* 2007). Regarding scale, for example, in data collected from within a single 10 km square, large-scale autocorrelation would result in patterns that show patches

of similar values over a kilometre or more, whilst data showing fine-scale autocorrelation may show similarity only over much smaller distances. In terms of strength, observations from patterns with weak autocorrelation will show considerable variation even over short distances, whilst patterns with strong spatial autocorrelation should lead to data with only small differences between neighbouring points. Whether autocorrelation is locally weak or strong, it can decay with distance quickly or instead be relatively persistent (Fig. 1a,b,d,e).

Whatever its nature, spatial autocorrelation does not in itself cause problems for analysis in the event that (1) the extrinsic causes of spatial pattern of the y variable are fully accounted for by the spatial structure of the measured x variables (i.e. all the systematic autocorrelation in the dependent variable is a simple function of the autocorrelation in the explanatory variables), and (2) intrinsic causes of spatial autocorrelation in the response (such as dispersal)

are absent (Cliff & Ord 1981). If both conditions are met, the errors about the regression model are expected to have no spatial autocorrelation and thus do not violate the assumptions of standard regression methodologies. In practice, the two conditions are almost never met simultaneously as, firstly, we can never be sure of including all the relevant x variables and, secondly, dispersal is universal in ecology. In this case, the errors are expected to be spatially dependent, violating an important assumption of most basic statistical methods. It is this spatial autocorrelation in the errors that, if not explicitly and correctly modelled, has a detrimental effect on statistical inference (Legendre 1993; Lichstein *et al.* 2002; Zhang *et al.* 2005; Barry & Elith 2006; Segurado *et al.* 2006; Beale *et al.* 2007; Dormann *et al.* 2007). In short, ignoring spatial autocorrelation in the error term runs the risk of violating the usual assumption of independence: it produces a form of pseudoreplication (Hurlbert 1984; Haining 1990; Cressie 1993; Legendre 1993;

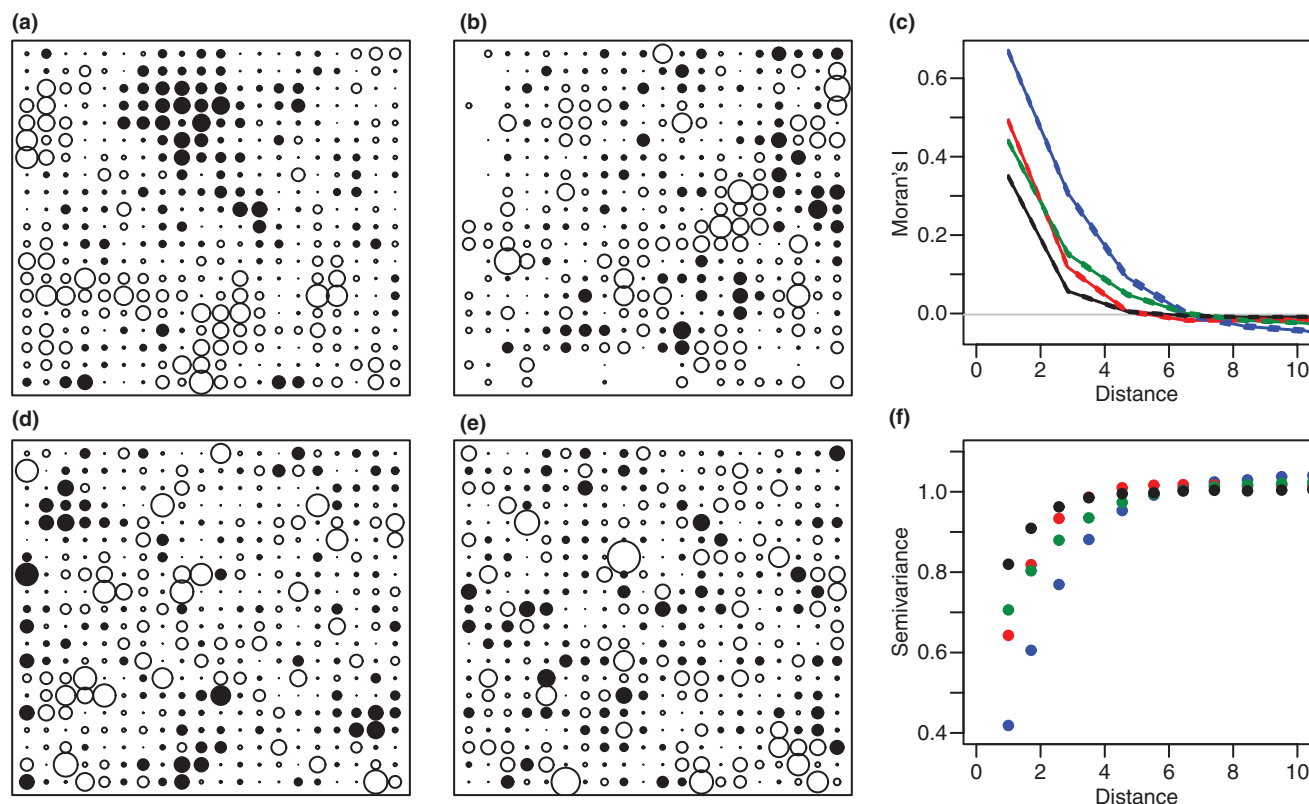


Figure 1 Examples of data sets showing spatial autocorrelation of both different scales and strengths and some basic exploratory data analysis. In (a), (b), (d) and (e), point size indicates parameter values, negative values are open and positive values are filled. (a) Large scale, strong autocorrelation; (b) large scale, weak autocorrelation; (d) small scale, strong autocorrelation; (e) small scale, weak autocorrelation. Correlograms (c) and empirical semi-variograms (f) showing mean and standard errors from 100 simulated patterns with (blue) large scale, strong autocorrelation, (green) large scale, weak autocorrelation, (black) small scale, weak autocorrelation and (red) small scale, strong autocorrelation. The expected value of Moran's I in the absence of autocorrelation is marked in grey. Note that correlograms for simulations with the same scale of autocorrelation cross the expected line at the same distance, and strength of autocorrelation is shown by the height of the curve.

Fortin & Dale 2005). Unsurprisingly, spatial pseudoreplication increases the Type I statistical error rate (the probability of rejecting the null hypothesis when it is in fact true) in just the same way as do other forms of pseudoreplication: *P*-values from non-spatial methods applied to spatially autocorrelated data will tend to be artificially small and so model selection algorithms will tend to accept too many covariates into the model (Legendre 1993; Lennon 2000; Dale & Fortin 2002, 2009; Barry & Elith 2006). This effect is easily illustrated by simulation (Fig 2a), showing that Type I errors from a non-spatial regression method (ordinary least squares, OLS) increase dramatically with the degree of autocorrelation in the errors, whilst those from a spatial regression method which correctly models the autocorrelation (generalized least squares, GLS) do not.

A related phenomenon is perhaps less well known: when the correct covariates are included in the model, the estimates of regression coefficients from methods which incorrectly specify the correlation in the errors are less precise (Cressie 1993; Fortin & Dale 2005; Beale *et al.* 2007). If the true regression coefficients are close to zero, then a decrease in estimation precision will lead to an increased chance of obtaining an estimate with a larger absolute value (Beale *et al.* 2007). Comparisons of the distributions of parameter estimates from application of Ordinary Least Squares and Generalized Least Squares to simulated data sets show this clearly, with strengthening autocorrelation resulting in an increasing tendency for Ordinary Least Squares estimates to be larger in magnitude than Generalized Least Squares estimates (Fig. 2b,c,d). When spatial autocorrelation in the errors is absent, the two methods are

broadly in agreement, but as either strength or scale of spatial autocorrelation increases, Ordinary Least Squares estimates become much more widely spread than Generalized Least Squares estimates. Whilst there is a mathematical proof for the optimal performance of Generalized Least Squares estimation when the correlation matrix of the errors is known (Aitken 1935), its performance in practice depends on the quality of estimation of the correlation matrix. The mathematical intractability of this problem has led to it being investigated by simulation (Alpargu & Dutilleul 2003; Ayinde 2007). This reduction in precision is probably responsible for the evidence in the ecology literature (Dormann 2007) that parameter estimates from spatially explicit modelling methods are usually of smaller magnitude than those from non-spatially explicit models applied to the same data sets. This also explains the unpredictability of the difference between regression coefficients from spatial and non-spatial methods (Bini *et al.* 2009); by using the very low precision estimate from ordinary least squares as the gold-standard against which other estimated regressions coefficients are judged, this study necessarily generates unpredictable differences.

EVALUATION AND SYNTHESIS OF SPATIAL REGRESSION METHODS

Data set scenarios

Simulation studies of spatial methods have been undertaken before in ecology (e.g. Beale *et al.* 2007; Dormann *et al.* 2007; Kissling & Carl 2008), but have been both insuffi-

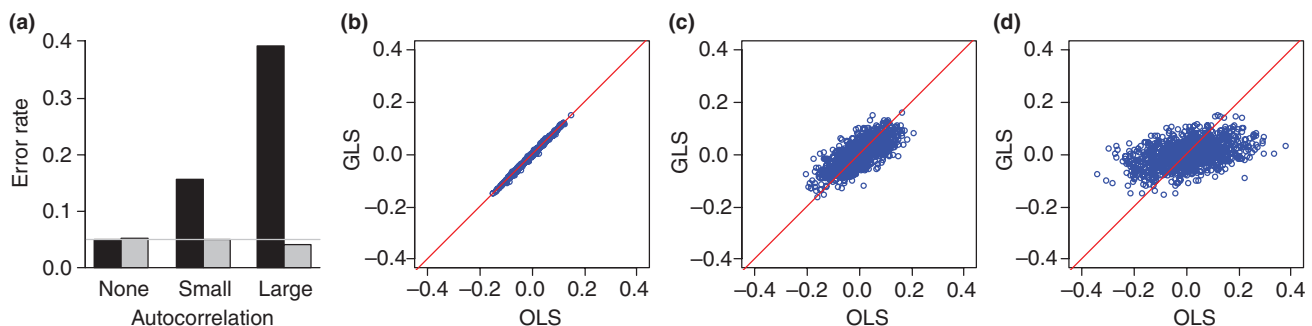


Figure 2 Two important consequences of spatial autocorrelation for statistical modelling. (a) Type I statistical error rates for the correlation between 1000 simulations of two independent but spatially autocorrelated variables estimated using Ordinary Least Squares (black) and Generalized Least Squares (grey) methods with increasing scale of autocorrelation ($0 < \alpha$; 2 grid squares $< \alpha$; 5 grid squares). Note that Type I error rates for models fitted using Ordinary Least Squares increase with scale of autocorrelation and are far greater than the nominal 0.05. Comparison of parameter estimates for the relationship between 1000 simulations of two independent but spatially autocorrelated variables with increasing scale of autocorrelation [same datasets as in (a)] is shown in (b–d). The Ordinary Least Squares and Generalized Least Squares parameter estimates are nearly identical in the absence of autocorrelation (b) but estimates from Ordinary Least Squares become significantly less precise as autocorrelation increases (c, d) whilst the distribution of estimates from models fitted with Generalized Least Squares are less strongly affected. Consequently, parameter estimates from model fitted with Generalized Least Squares are likely to be smaller in absolute magnitude than those from Ordinary Least Squares methods. The simulated errors were normally distributed and decayed exponentially with distance, whilst the Generalized Least Squares method used a spherical model for residual spatial autocorrelation.

ciently broad and too simplistic to reflect the complexities of real ecological data (Diniz-Filho *et al.* 2007). Here, we simulate spatial data sets covering eight scenarios that reflect much more the complexity of real ecological data sets. Full details of the implementation and R code for replicating the analysis are provided as Supporting Information, here (to maintain readability) we present only the outline and rationale of the simulation process. The basis for each scenario was similar: we simulated 1000 data sets of 400 observations on a 20×20 regular lattice. To construct each dependent variable, we simulated values for the covariates using a Gaussian random field with exponential spatial covariance model. All scenarios incorporated pairs of autocorrelated covariates that are cross-correlated with each other (highly cross-correlated spatial variables cause imprecise parameter estimates in spatial regression: see Supporting Information). We then calculated the expected value for the response as a linear combination (using our chosen values for the regression coefficients) of the covariates, then simulated and added the (spatial) error term as another (correlated) Gaussian surface. Just as with real ecological data sets, all our scenarios include variables which vary in both the strength and the scale of autocorrelation. As strong or weak autocorrelation are relative terms, here we defined weakly autocorrelated variables as having a nugget effect that accounts for approximately half the total variance in the variable, and strongly autocorrelated patterns as having a negligible nugget. Similarly, large- and small-scale autocorrelation is relative, so here we define large-scale autocorrelated patterns as having an expected range approximately half that of the simulated grid (i.e. 10 squares), whilst small scale had a range of around one-third of this distance.

Within this basic framework, scenarios 1–4 (referred to below as ‘simpler scenarios’) involve covariates and correlation matrices for the errors which are homogeneous, meaning that the rules under which the data are generated are constant across space and do not violate the homogeneity assumption made by many spatial regression methods. Scenarios 5–8 (described below as ‘more complex’) all involve adding an element of spatial inhomogeneity (i.e. non-stationarity) to the basic situation described above and therefore at least potentially violate the assumptions of all the regression methods we assessed (Table 1). We note that non-stationarity is used to describe many different forms of inhomogeneity, and here we incorporated non-stationarity in several different ways: firstly, we included spatial variation in the true regression coefficient between the covariates and the dependent variable (smoothly transitioning from no relationship – i.e. regression coefficient = 0 – along one edge of the simulated surface, to a regression coefficient of 0.5 along the opposite edge). Secondly, we included covariates with non-stationary autocorrelation structure, implemented such that one edge of the simulated surface had a large-scale autocor-

relation structure gradually changing to another with small-scale autocorrelation structure (as commonly seen in real environmental variables such as altitude when a study area includes a plain and more topographically varied area). Thirdly, we incorporated a spatial trend in the mean: another form of non-stationarity. And finally, we incorporated similar types of non-stationarity in the mean and/or correlation structure of the simulated errors.

We simulated datasets with exponential autocorrelation structures because our method for generating cross-correlated spatial patterns necessarily generates variables with this structure, although alternative structures are available if cross-correlation is not required. In the real world, environmental variables exhibit a wide range of spatial autocorrelation structures.

For each scenario, we estimated the regression coefficients using all the methods listed below (Table 2). We then summarized the coefficient estimates (excluding the intercept) for each statistical method, assessing performance in terms of precision and bias. Contrary to standard definitions of precision which measure spread around the mean of the parameter estimates, here we measure mean absolute difference from the correct parameter estimate; a more meaningful index for our purposes. For methods where selection of covariates is possible, we also record the Type I and Type II statistical error rates.

For each method of estimation and each scenario, performance statistics were evaluated in the form of the median estimate of absolute bias and root mean square error (RMSE, the square root of the mean squared difference between the estimates and the associated true values underlying the simulated data). These were then combined across scenarios after rescaling by the corresponding values for Generalized Least Squares-Tb (Table 2).

Model fitting and parameter estimation

A wide range of statistical methods have been used in the literature for fitting regression models to spatial data sets (Table 2, where full details and references can be found for each method), and a number of recent reviews have each highlighted some methods whilst explicitly avoiding recommendations (Guisan & Thuiller 2005; Zhang *et al.* 2005; Barry & Elith 2006; Elith *et al.* 2006; Kent *et al.* 2006; Dormann *et al.* 2007; Miller *et al.* 2007; Bini *et al.* 2009). Each review concludes that different methods applied to identical data sets can result in different sets of covariates being selected as important, due to the many differences underlying the methods (e.g. in modelled correlation structures and computational implementation).

Methods for fitting linear regression models can be classified according to the way spatial effects are included (Dormann *et al.* 2007). Three main categories exist: (i)

Table 1 Scenarios for assessing performance of statistical methods applied to spatial data. In all scenarios, the covariates and error term have an exponential structure underlying any added non-stationarity. R code provided in the Supporting Information provides a complete description of all scenarios, Figs S4–S10 identify the correlated variables and the expected value of each parameter

Scenario	Designed to test	Dependent variable error	Covariates
1	The performance of models when assumptions are met, but with correlated x variables which also have various strengths and scales of autocorrelation	Strong (all variance because of spatial pattern), large-scale (≈ 10 grid squares) autocorrelation	Six x variables having varying scales and strengths of autocorrelation, with a subset correlated (expected Pearson's correlation = 0.6) with each other. Four have non-zero regression coefficients in the simulation of the y -variable
2	As (1)	Strong, small-scale (≈ 2 grid squares) autocorrelation	As (1)
3	As (1)	Weak (50% of variance because of spatial structure), large-scale autocorrelation	As (1)
4	As (1)	Weak, small-scale autocorrelation	As (1)
5	The performance of methods when x variables have various kinds of non-stationarity	Strong, large-scale autocorrelation	Three x variables, one of which also has non-stationary autocorrelation structure. Two have non-stationary correlations with each other. The third x variable has intermediate scale autocorrelation (≈ 5 grid squares) and a strong (i.e. adding equal variance to the pattern) linear spatial trend
6	As (5)	Strong, small-scale autocorrelation	As (5)
7	The performance of methods when the errors in the y variable are non-stationary in scale of autocorrelation	Non-stationary: varying large to small scale autocorrelation across domain	As (5) but all variables are uncorrelated
8	The performance of methods when the errors in the dependent variable have a trend	Non-stationary: strong large-scale autocorrelation plus strong (i.e. adding equal variance to pattern) trend	As (7)

methods that model spatial effects within an error term (e.g. Generalized Least Squares, implemented here using a Spherical function for the correlation matrix of the errors (GLS-S), a structure that is deliberately different to the exponential structure of the simulated data and Simultaneous Autoregressive Models (SAR), implemented here as an error scheme, which is Generalized Least Squares with a 1-parameter model for the correlation matrix of the errors), (ii) methods incorporating spatial effects as covariates (e.g. Spatial Filters and Generalized Additive Models) and (iii) methods that pre-whiten the data, effectively replacing the response data and covariates with the alternative values that are intended to give independent values for analysis (e.g. Wavelet Revised Models). From these three categories, we selected a total of 11 different methods (with 10 additional variants, including Generalized Additive Mixed Models which allows for a spatial term in both the covariate and

error terms), covering the range used in the ecological and statistical literature (Table 2).

Where relevant, we specified a spherical covariance structure for the errors during parameter estimation rather than the correct exponential structure because in real-world problems the true error structure is unknown and is unlikely to exactly match the specified function. Consequently, it is important to know how these methods perform when the error structure is not modelled exactly to assess likely performance in practical situations. Although several techniques have been used for selecting spatial filters (Bini *et al.* 2009), only one method – the selection of filters with a significant correlation with the response variable – has been justified statistically (Bellier *et al.* 2007) and consequently we use this implementation. We also include two generalized least square models we call GLS-True that had the correct empirical spatial error structure. These models are

Table 2 Spatial analysis tools applied to each of the 1000 simulations of eight scenarios

Method	Description	Classification
Ordinary Least Squares (OLS)	Most basic regression analysis, regarding errors about the fitted line as being independent and with equal variance	Non-spatial
OLS with model selection (OLS MS)	As OLS with stepwise backward elimination of non-significant variables using <i>F</i> -tests and 5% significance	Non-spatial
Subsampling (SUB) (Hawkins <i>et al.</i> 2007)	Data set repeatedly resampled at a scale where no autocorrelation is detected, OLS model fitted to data subsets and mean parameter estimates from 500 resamples treated as estimates	Non-spatial
Spatial Filters (FIL)(Bellier <i>et al.</i> 2007)	A selection of eigenvectors (those significantly correlated with the dependent variable) from a principal coordinates analysis of a matrix describing whether or not locations are neighbours are fitted as nuisance variables in an OLS framework	Space in covariates
FIL with model selection (FIL MS)	As FIL, with stepwise backward elimination of non-significant covariates (but maintaining all original eigenvectors) using <i>F</i> -tests and 5% significance.	Space in covariates
Generalized Additive Models with model selection (Generalized Additive Models MS)	As Generalized Additive Models but stepwise backwards elimination of non-significant covariates using χ^2 tests and 5% significance (degrees of freedom in thin plate spline fixed as that identified before model selection)	Space in covariates
Simple Autoregressive (AR) (Augustin <i>et al.</i> 1996; Betts <i>et al.</i> 2009)	An additional covariate is generated consisting of an inverse distance weighted mean of the dependent variable within the distance over which spatial autocorrelation is detected. Ordinary Least Squares is then used to fit the model	Space in covariates
AR with model selection (AR MS)	As AR but stepwise backwards elimination of non-significant covariates using <i>F</i> -tests and 5% significance	Space in covariates
Wavelet Revised Models (WRM) (Carl & Kuhn 2008)	Wavelet transforms are applied to the covariate matrix and the transformed data analysed using Ordinary Least Squares	Spatial correlation removed from response variable, with corresponding redefinition of the covariates
Simultaneous Autoregressive (SAR) (Lichstein <i>et al.</i> 2002; Austin 2007; Kissling & Carl 2008)	Spatial error term is predefined from a neighbourhood matrix and autocorrelation in the dependent variable estimated, then parameters are estimated using a GLS framework. Here, we use simultaneous autoregressive error models with all first order neighbours with equal weighting of all neighbours (Kissling & Carl 2008)	Space in errors
SAR with model selection (SAR MS)	As SAR but stepwise backwards elimination of non-significant covariates using likelihood ratio tests and 5% significance	Space in errors
Generalized Additive Mixed Models (GAMM) (e.g. Wood 2006)	An extension of GAM to include autocorrelation in the residuals.	Space in errors and covariates.
Generalized Additive Mixed Models with model selection (GAMM MS)	As GAMM but stepwise backwards elimination of non-significant covariates using likelihood ratio tests and 5% significance	Space in errors and covariates
Generalized Least Squares (GLS-S) (Pinheiro & Bates 2000)	A standard generalized least squares analysis, fitting a spherical model of the semi-variogram. Model fitting with REML	Space in errors
GLS with model selection (GLS-S MS)	As GLS but model fitting with ML and stepwise backwards elimination of non-significant covariates using likelihood ratio tests and 5% significance	Space in errors

Table 2 *continued*

Method	Description	Classification
Bayesian Conditional Autoregressive (BCA) (Besag <i>et al.</i> 1991)	A Bayesian intrinsic conditional autoregressive (CAR) model using all first order neighbours with equal weighting and analysed via MCMC using the WinBUGS software (Lunn <i>et al.</i> 2000) with 10000 iterations for each analysis	Space in errors
BCA with model selection (BCA MS)	As above but using reversible jump variable selection (Lunn <i>et al.</i> 2006), selecting the model with highest posterior probability	Space in errors
True GLS (GLS T)	As GLS but spatial covariance defined and fixed <i>a priori</i> from the exponential semi-variogram of the actual (known) error structure. For scenarios 1–6, this reflects the true structure of the simulations in a way that is not possible in real data	Space in errors
True GLS with model selection (GLS T MS)	As true GLS but model fitting with ML and stepwise backwards elimination of non-significant covariates using likelihood ratio tests and 5% significance	Space in errors
True GLS b (GLS Tb)	As GLS T but with only the correct covariates included within the initial model	Space in errors
True GLS b with model selection (GLS Tb MS)	As GLS T MS but with only the correct covariates included within the initial model	Space in errors

impossible in real-world analysis but provide an objective best-case gold-standard to measure other parameter estimates against in addition to the true parameter estimates and are therefore repeatable comparisons as further spatial regression methods are developed.

All these methods, and the data simulation, have been implemented using the free software packages R (R development core team 2006) and, for the conditional autoregressive model, WinBUGS (Spiegelhalter *et al.* 2000). To facilitate the use of our scenarios and provide a template for ecologists interested in undertaking their own spatial analysis, all the code to generate the simulations and figures presented here is provided as Supporting Information.

The distinctions between the various regression models are extremely important in terms of interpretation of the results, as the expected patterns of residual autocorrelation vary between the three categories. Contrary to assertions by some authors (Zhang *et al.* 2005; Barry & Elith 2006; Segurado *et al.* 2006; Dormann *et al.* 2007; Hawkins *et al.* 2007), the residuals of a correctly fitted spatial model may not necessarily lack autocorrelation. Take the case of two spatially autocorrelated variables (y and one x) that in truth are independent of each other (Fig. 3). In this simple example, all methods should, on average, correctly estimate the slope to be zero. However, models that ignore spatial effects will clearly have autocorrelation in the residuals, violating the model assumptions and resulting in lower precision and inflated Type I error rates. By contrast, those models that assign spatial effects to an error term will also retain autocorrelation in the residuals because the error term forms part of the residual variation (i.e. variation that

remains after the covariate effects – in this case expected to be zero – are accounted for; residuals and errors are synonyms in this usage) but the important difference is that these models are tolerant of such autocorrelation and should provide precise estimates and correct error rates. When fitted correctly, the third class of models with spatial processes incorporated in the fixed effects should show little residual autocorrelation, even when (as in this simple example) the spatial structure in the dependent variable is entirely unrelated to the covariate, because such structure should be ‘soaked up’ by the additional covariates.

SPATIAL ANALYSIS METHOD PERFORMANCE

The performance of the different methods is summarized in Table 3 and all results are presented graphically in Figs S4–S11, with an example (Scenario 1) illustrated in Fig. 4. In fact, the best performing methods in any one scenario also tended to be the best performing methods in other scenarios.

Focussing first on the four simple scenarios (Figs S4–S7, Fig. 4), in the absence of model selection, all the methods with autocorrelation incorporated in the error structure perform approximately equally well in terms of absolute bias and root mean squared error. Methods incorporating spatial structure within the covariates were generally much poorer, with the exception of the Generalized Additive Models methods which were only marginally poorer. With the exception of Generalized Additive Models and Wavelet Revised Models, methods that did not have space in the errors had the greatest difficulty estimating parameters for

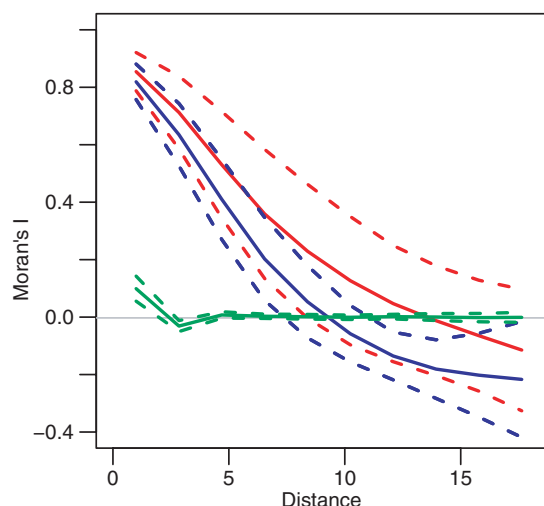


Figure 3 Autocorrelation of the residuals from two types of spatial analysis. Generalized Least Squares methods (red) have autocorrelation modelled in a spatial error term (here as an exponential structure) so errors are correctly autocorrelated, whilst wavelet revised methods (green) have autocorrelation removed before analysis. The plot shows mean and standard deviations from 100 simulations of Moran's I for residuals. The mean and standard error of the autocorrelation structure of the dependent variables is plotted in blue: note the similar autocorrelation structure of the residuals from Generalized Least Squares models and the dependent variable, but that the residuals of the Wavelet Revised Models are effectively zero after the first lag. Both models are correctly fitted, despite the Generalized Least Squares fit retaining residual autocorrelation.

cross-correlated covariates. Subsampling to remove autocorrelation (SUB) was consistently the worst method. Ordinary Least Squares performed poorly in the presence of strong, large-scale autocorrelation, but before model selection was otherwise comparable with the other unbiased methods. Applying model selection to the well-performing methods resulted in a consistent and marked improvement, but model selection with other methods (including Ordinary Least Squares) resulted in less consistent improvement in their performance and sometimes in no improvement whatsoever. Regarding statistical errors, all methods showed low Type II error rates (failure to identify as significant the covariates whose regression coefficients were in truth not zero). Ordinary Least Squares and Simple Autoregressive showed particularly high Type I error rates (identifying as significant covariates whose regression coefficients were in truth zero). Type I error rates were generally above the nominal 5% rate, but lower for methods with autocorrelation in the error structure: Simultaneous Autoregressive Models performed well for three of the four scenarios, whereas Generalized Least Squares-S performed relatively poorly for three of the four scenarios. For Bayesian

Conditional Autoregressive, the error rate was consistently under half of the (otherwise) nominal 5% level. Generalized Least Squares-T, which had almost exactly the correct Type I error rate for three scenarios, had double the nominal value for the third scenario.

The first two of the more complex scenarios where non-stationarity was introduced in the covariates (Figs S8 and S9), and hence indirectly into the residuals, generally produced a marked decline of the performance of poorer methods (from the simple scenarios) against that of the better methods. Good methods were again those with space in the error terms and Generalized Additive Models. Model selection, whilst having little effect on absolute bias, improved the precision of estimates from the good methods (i.e. those with space in the errors and Generalized Additive Models) but not those of the poorer methods (the remaining methods). In particular, spatial filters and autoregressive (AR) methods were highly biased and subsampling again resulted in imprecise estimates. Increased scale of autocorrelation in the errors in the y variable again resulted in Ordinary Least Squares performing more poorly. Ordinary Least Squares, Simple Autoregressive and Spatial Filters methods had Type I error rates of 100%, Generalized Least Squares-S had the highest rate of the better performing methods, while those of Bayesian Conditional Autoregressive and Simultaneous Autoregressive Models were close to target. Type II error rates were generally lower apart from Ordinary Least Squares, Spatial Filters and Simple Autoregressive Models.

The last two scenarios (Figs S10 and S11), with non-stationarity introduced in the error term (and therefore the most challenging), generated a comprehensive failure for most methods: only Generalized Least Squares when provided with the correct model parameters for the autocorrelation of the errors and the correct set of covariates beforehand (Generalized Least Squares-Tb) performed very well. This is unsurprising given that it was provided with information that would be unknown in most circumstances. The distinction between the good and bad methods was still evident even in these extreme scenarios (Scenario 8), but only if autocorrelation was not both strong and large scale (Scenario 7). The Generalized Additive Mixed Models methods proved impossible to fit irrespective of autocorrelation. Simple Generalized Additive Models performed best of all, particularly with model selection. In Scenario 7, Ordinary Least Squares, Simple Autoregressive and Spatial Filters always found covariates significant when there was no true relationship. Other methods also had inflated Type I error rates, but were broadly comparable. Simple Autoregressive and Spatial Filters had high Type II error rates. In Scenario 8, Generalized Additive Mixed Models methods again failed to fit the simulated data, whilst Type I errors were uniformly too high (except for

Table 3 Overall performance of spatial analysis methods

Method	Overall performance	Comments
OLS	Moderate precision and bias, but imprecise with strong large-scale AC	
OLS MS	Highly biased, intermediate to poor Type I and poor Type II error rate	
OLSGLS	Highly biased, intermediate to poor Type I and poor Type II error rate	c. 10% of simulations failed to converge
SUB	Extremely low precision	
FIL	Highly (downward) biased	Moderately computer intensive
Generalized Additive Models MS	Fairly good overall performance. Intermediate Type I error rate throughout, poor Type II error rate	Convergence issues for only one model
AR	Highly (downward) biased	
AR MS	Highly (downward) biased, Type I error rates were good for scenarios 1–4, poor for 5–8. Type II error rates were intermediate to poor	
SAR	Generally good overall performance	
Generalized Additive Models Generalized Additive Mixed Models MS	Generally good overall performance. Type I error rates were good to intermediate, Type II error rates were good throughout	Model convergence was never achieved for scenarios 7 and 8 and frequently failed in other scenarios. Extremely computationally intensive
GLS-S	Generally good overall performance	Convergence issues for several simulations in scenarios 1, 5, 6, 7 and 8. Computationally intensive
GLS-S MS	Generally good overall performance. Intermediate Type I error rate throughout, poor Type II error rate	Convergence issues for several simulations in scenarios 1, 5, 6, 7 and 8. Very computationally intensive
BCA	Good overall performance	Moderately computationally intensive. Improvements would be made by manual checking and tuning of MCMC chains
BCA MS	Best overall performance	Moderately computationally intensive. Improvements would be made by manual checking and tuning of MCMC chains
GLS T	Good overall performance	Not possible with real data
GLS T MS	Very good overall performance	Not possible with real data
GLS Tb	Excellent overall performance	Not possible with real data, but demonstrates the value of <i>a priori</i> knowledge
GLS Tb MS	Excellent overall performance	Not possible with real data

Generalized Least Squares-T, which was too low). As all methods identified the vast majority of covariates as significant irrespective of true effect, Type II error rates were low, and some (Generalized Least Squares-T, Generalized Least Squares-S, Simultaneous Autoregressive Models

and Spatial Filters) were effectively zero. Ordinary Least Squares, Simple Autoregressive, Spatial Filters and Simultaneous Autoregressive Models had high Type I error rates, that for Generalized Least Squares-S was nearly as bad and only Bayesian Conditional Autoregressive models and

Generalized Additive Models were better, but still poor. High Type II errors were encountered with both Ordinary Least Squares and Simple Autoregressive, with BCA and Generalized Additive Models intermediate.

Considering the performance of the models within scenarios, and also in combination across all scenarios (Fig. 5), the following results can be observed.

- (1) Differentials in method performance in scenarios where model assumptions were not violated (the 'simpler scenarios') can be seen to have been exaggerated in the more complex scenarios. Poorly performing methods in the simpler scenarios performed much worse in the complex scenarios.
- (2) Applying model selection using methods with low bias generally resulted in improved precision and lower bias (cf. Whittingham *et al.* 2006). In contrast, applying model selection to Ordinary Least Squares models resulted in less substantial change in model performance – this is notably the case in Scenarios 5 and 6. We note also that the three-stage process of initially using Ordinary Least Squares methods, applying model selection and then using a Generalized Least Squares-S method on the significant variables did not reliably improve on the performance of the Ordinary Least Squares methods in the more complex scenarios.
- (3) Methods that modelled space in the residuals always had lower Type I error rates than Ordinary Least Squares. The 'true Generalized Least Squares models' – that is, models fitted by Generalized Least Squares where the autocorrelation structure is set to the actual structure used in the simulations rather than being estimated from the data – have Type I error rates close to 5% for four of the first six scenarios. In these ideal case methods such as Generalized Least Squares where correlation structures can be fixed are consistently the most accurate, but clearly this can never be applied in real situations and the increased Type I error rate associated with incorrect specification of the error structure is evident from our simulations.

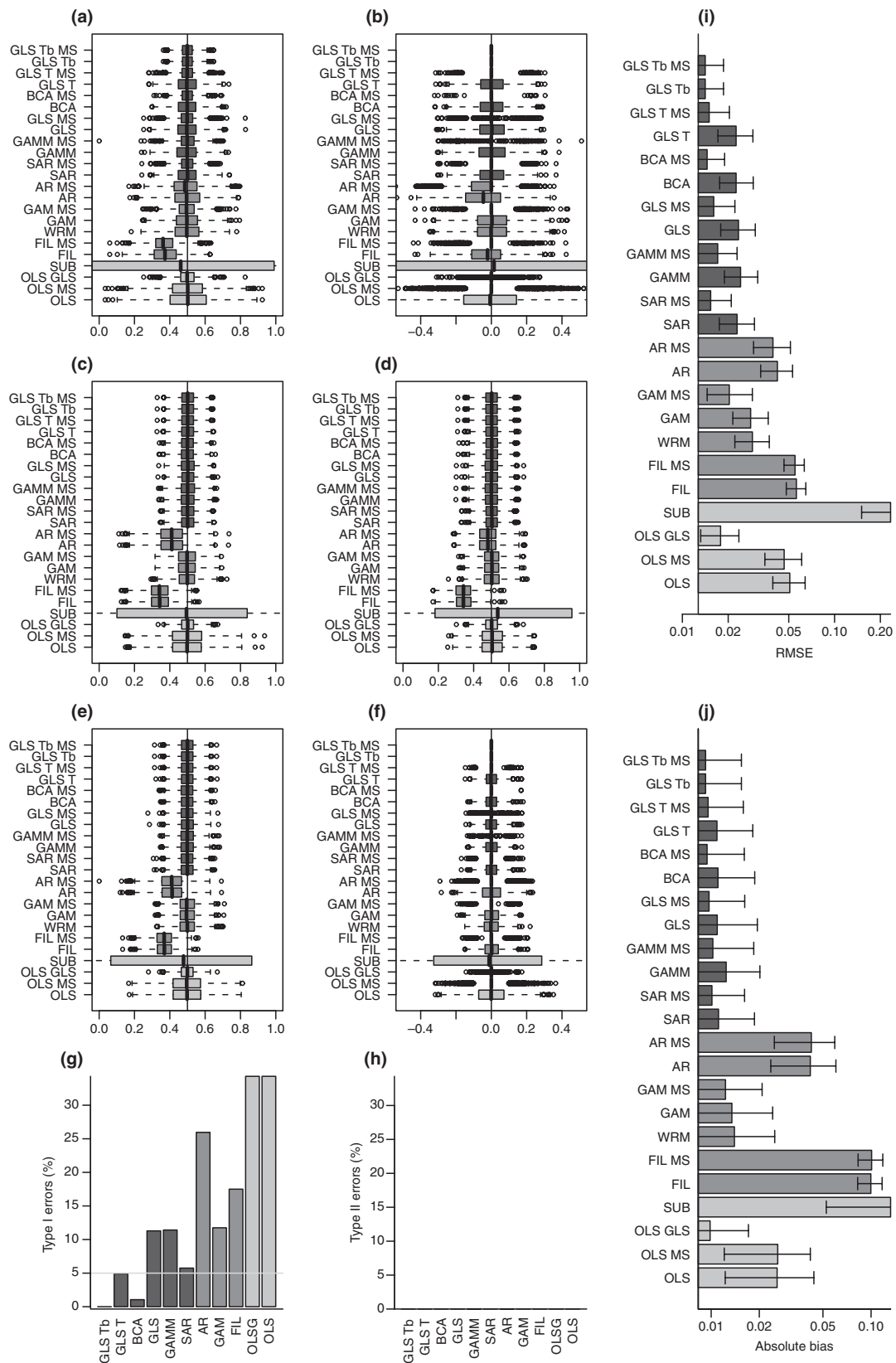
- (4) As described by others (e.g. Burnham & Anderson 2003), including only the correct covariates consistently resulted in better parameter estimates for the remaining parameters: it is not enough to rely on the data to give an answer if *a priori* knowledge of likely factors is available.
- (5) With the exception of scenario 8, spatial filters and simple autoregressive models were generally poor.

In summary, there appear to be a suite of methods giving comparable absolute bias and RMSE performance measures in the absence of model selection for the first six scenarios, including Generalized Least Squares-S, BCA, Simultaneous Autoregressive Models Generalized Additive Mixed Models and Generalized Additive Models, with Ordinary Least Squares performing rather less well. Relative to these performance measures, performance of all these methods was generally improved by model selection, the gains made by Ordinary Least Squares being least marked due to high Type I error rates. Generalized Additive Models and Generalized Least Squares-S had the next highest Type I error rates, those of Simultaneous Autoregressive Models and (where it converged) Generalized Additive Mixed Models were on average close to 5%, whilst that of BCA, which had the lowest values for comparable absolute bias and RMSE after model selection, lay between 1 and 2%.

Data partitioning and detrending

One potential problem with spatial analysis is that model fitting can involve unreasonably large computation times. The main reason for this is that for some methods, the computation time depends on the number of possible pairwise interactions between points. For such methods, one way of dealing with the combinatorial problem is to split the data and analyse the subsets (Davies *et al.* 2007; Gaston *et al.* 2007). We consider two ways of splitting a data set: (1) random partition (randomly allocating each square to one of two equal size groups); and (2) simple blocking via two contiguous blocks. We can then fit spatial models and

Figure 4 An example showing the results of models (described in Table 2) fitted to 1000 simulations of Scenario 1 (described in Table 1). Plots (a–f) illustrate boxplots of the 1000 parameter estimates for the 1000 realizations of Scenario 1 for each of the different modelling methods with the true parameter value for each covariate indicated by the vertical line [true value = 0 for (b and f), 0.5 for (a, c, d and e)]. Panels (b–f) are the parameter estimates for the six covariates with varying scales and strength of autocorrelation and correlations among each other. All covariates are autocorrelated, cross-correlations exist between two pairs of covariates (a) with (b) and (d) with (e). See main text and Supporting Information for further details. Type I and Type II error rates are illustrated in (g) and (h) respectively (NB there were essentially no Type II errors in this scenario, hence this figure appears empty). The average root mean squared error (i.e. difference between the parameter estimate and the true value) of all six parameter values in each of the 1000 realizations of the scenario (i) and the overall bias (i.e. systematic error from the true parameter value) (j) are also illustrated (note log scale). In each plot, non-spatial models are pale grey, spatial models with spatial effects modelled as covariates are intermediate shade and spatial models with spatial effects modelled in the errors are dark grey. Note that precision and bias are consistently low for spatial models with space in the errors and for these models model selection results in more accurate estimates. Abbreviations identify the modelling method and are explained in Table 2. A complete set of figures for the remaining seven scenarios are provided as Supporting Information.



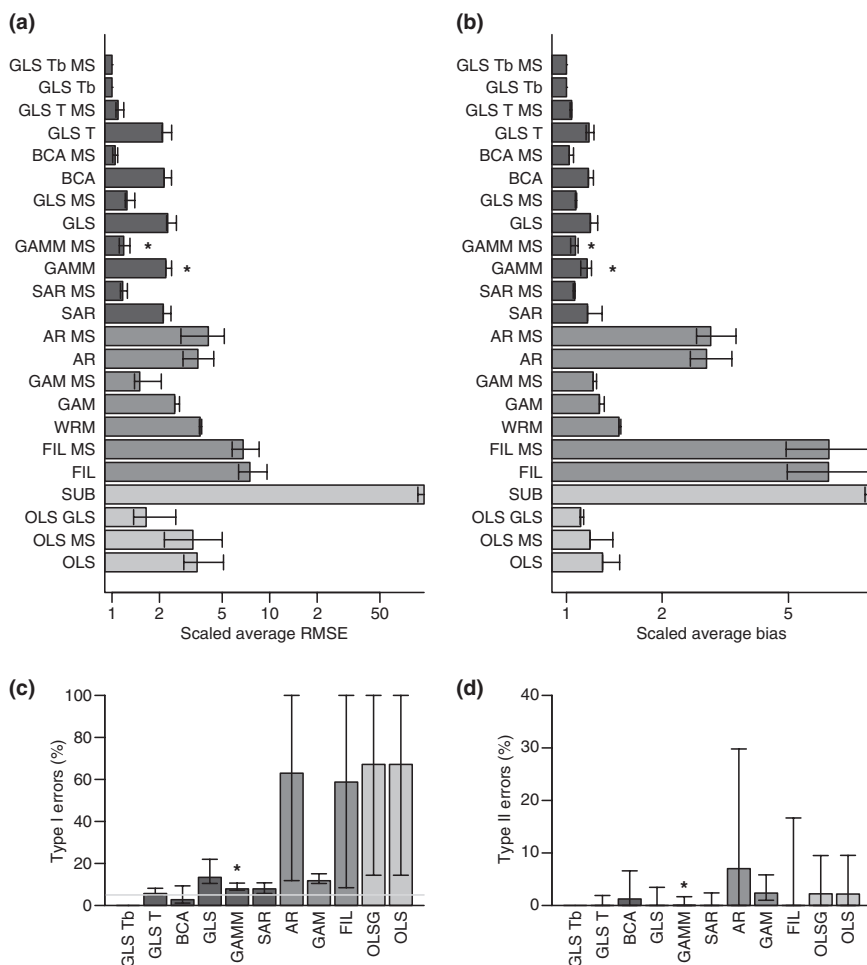


Figure 5 Relative precision (a), Relative bias (b), Type I statistical errors (c) and Type II statistical errors (d) across all eight simulation scenarios described in Table 1. In drawing this figure, the median values of absolute bias and RMSE for each method have been divided by the corresponding values for Generalized Least Squares-Tb before pooling across scenarios (hence the values are relative precision and bias, rather than absolute). This process ensured an even contribution from each scenario after standardization relative to a fully efficient method. Error plots show median and IQR. Asterisks indicate that Generalized Additive Mixed Models models never converged in Scenarios 7 and 8 (where other methods performed worst) so may be lower than expected.

compare bias and precision for the two partition methods. This shows that method (2) considerably reduces computation time (Fig. S12) at no cost to the precision of parameter estimation (Fig. S13). In contrast, method (1) results in far less precise parameter estimates because the number of cells in each group with very close neighbours in the same group (important for correct estimation of the covariance matrix) is much reduced in this case.

It is a common recommendation that any linear spatial trends identified in the dependent variable are removed before model fitting (Koenig 1999; Curriero *et al.* 2002). In fact, this is automated in some model fitting software (Bellier *et al.* 2007). To test the effects of this approach, we can remove linear trends in the dependent variable and again examine bias and precision. The result of detrending in this manner can be further compared with the effect of including x and y coordinates as additional covariates. We found that detrending results in significant bias towards parameter estimates of smaller magnitude (Fig. 6a) as it removes meaningful variation when the covariates also show linear trends. No such effect was observed when coordinates were included as additional covariates (Fig. 6b).

DISCUSSION

The central theme to be drawn from the results of our analysis comparing the various modelling methods is clear: some methods consistently outperform others. This reinforces the notion that ignoring spatial autocorrelation when analysing spatial data sets can give misleading results: in each scenario, the difference in precision between our best and worst performing methods is considerable. This overall result is completely in agreement with previous, less wide-ranging studies (e.g. Dormann *et al.* 2007; Carl & Kuhn 2008; Beguería & Pueyo 2009). Methods with space in the errors (Generalized Least Squares-S, Simultaneous Autoregressive Models, Generalized Additive Mixed Models, Bayesian Conditional Autoregressive) generally performed similarly and often considerably better than those with space in the covariates (Simple Autoregressive, Wavelet Revised Models, Spatial Filters) which are in turn generally better than non-spatial methods (particularly Subsampling). Introducing model selection improves most methods but still leaves the poorer methods lagging behind. For hypothesis testing, the statistical error rates are most important. In

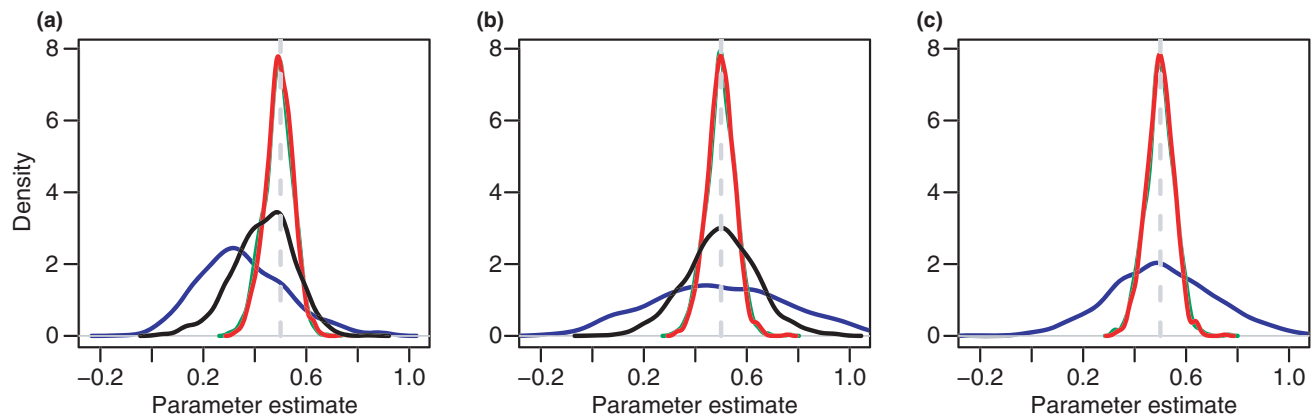


Figure 6 The effect on parameter estimates of removing linear spatial trends in the dependent variable (detrending) before statistical model fitting. All plots show parameters estimated from the same 1000 simulations, each with an autocorrelated x variable and autocorrelated errors in the response variable, where the true regression coefficient is 0.5 (dashed grey line). Ordinary Least Squares = blue, Wavelet Revised Models = black [not present in panel (c) where the model could not be fitted], Simultaneous Autoregressive Models = green (usually hidden under red line), Generalized Least Squares = red. In plot (a), the dependent variable was detrended before analysis, in plot (b) the uncorrected dependent variable was used and in plot (c) the x and y coordinates were included as covariates. Note both the clear bias in some parameter estimates and the skew caused by detrending [not evident in (c)], most evident in estimates from Ordinary Least Squares and Wavelet Revised Models methods but present regardless of model type.

general, all methods suffered from inflated Type I error rates in some scenarios, apart from Bayesian Conditional Autoregressive, which was consistently below the nominal 5% rate set for the other tests. The type I errors for Simultaneous Autoregressive Models, Generalized Least Squares-S, Generalized Additive Mixed Models and Generalized Additive Models were consistently better than those for Ordinary Least Squares, Spatial Filters and Simple Autoregressive. Note that the bias towards smaller parameter estimates in Spatial Filters and Simple Autoregressive methods is distinct from the observation that on a single data set spatial regression methods often result in smaller estimates than non-spatial regression methods (Dormann 2007). The latter is explained by the increased precision of spatial regression methods (Beale *et al.* 2007), the former is probably explained by both spatial filters and a locally smoothed dependent variable resulting in overfitting of the spatial autocorrelation, leaving relatively little meaningful variation to be explained by the true covariates. It is notable that this consistency of model differences suggests that, contrary to a recent suggestion otherwise (Bini *et al.* 2009), differences between regression coefficients from different models can be explained, but only when the true regression coefficient is known (Bini *et al.* 2009 assume regression coefficients from ordinary least squares form a gold-standard and measure the difference between this and estimates from other methods, a difference that depends mainly on the unknown precision of the least squares estimate, rather than the difference from the true regression coefficient which is of course unavailable in real data sets

examples). The particularly poor results for the subsampling methodology are entirely explained by the extremely low precision of this method caused by throwing away much useful information (see Beale *et al.* 2007 for a full discussion). Overall, we found Bayesian Conditional Autoregressive models and Simultaneous Autoregressive models to be among the best performing methods.

We analysed the effects of removing spatial trends in the y variable before analysis and the effect of splitting spatial data sets to reduce computation times. Our results do not support the removal of spatial trends in the y variable as a matter of course. If only local-scale variation is of interest it may be valid to include spatial coordinates as covariates within the full regression model provided there is evidence of broad global trends. We found that if large data sets are split spatially before analysis the accuracy and precision of parameter estimates are not unreasonably reduced.

Despite there being good reasons for anticipating autocorrelated errors in ecological data, it has often been suggested that testing residuals for spatial autocorrelation after ordinary regression is sufficient to establish model reliability. This is often carried out as a matter of course, with the assumption that if the residuals do not show significant autocorrelation the model results are reliable and vice versa (Zhang *et al.* 2005; Barry & Elith 2006; Segurado *et al.* 2006; Dormann *et al.* 2007; Hawkins *et al.* 2007). However, caution is required here, as the failure to demonstrate a statistically significant spatial signature in the residuals does not demonstrate absence of potentially influential autocorrelation. As our simulation study has

demonstrated, even weak autocorrelation can have dramatic effects on parameter estimation, so it is unwise to rely solely on this type of *post hoc* test when assessing model fit. Moreover, correctly fitted spatially explicit models will often show autocorrelated residuals (Fig. 3) so this test should certainly not be considered as identifying a problem with such models.

Whilst we designed our simulations to explore a wide range of complexity likely in real data, they do not cover all possible complications. In particular, all our simulated covariates and residuals have an exponential structure as a consequence of needing to simulate cross-correlated variables, yet real-world environmental variables may have a range of different autocorrelation structures. Although empirically our results are limited to the cases we considered, we found consistent patterns that are likely to remain generally true: we found that the differences in model performance in simple scenarios were only exaggerated in the more complex scenarios where a number of model assumptions were violated. Once suitable simulation methods are developed, future work could usefully explore alternative autocorrelation structures and confirm that similar results are also found under these conditions. Compared with previous, more restricted simulation studies (e.g. Dormann *et al.* 2007; Carl & Kuhn 2008; Kissling & Carl 2008; Beguería & Pueyo 2009), we find the same overall result – that spatially explicit methods outperform non-spatial methods – but our results show that the differences between modelling methods when faced with assumption violations and cross-correlation are less distinct; several methods that modelled space in the errors were more or less equally good.

Faced with a real data set, it is difficult to determine *a priori* which of the scenarios we simulated are most similar to the real data set. As we found that *all* the methods applied to certain scenarios (e.g. Scenario 8) were very misleading, it is important to determine whether the fitted model is appropriate and, if necessary, try fitting alternative model structures; in the case of Scenario 8, including spatial coordinates as covariates would dramatically improve model performance. This is a very important point: if model assumptions are badly violated, no regression method, spatial or non-spatial, will perform well, no matter how sophisticated. It is therefore vital that, within the limitations of any data set, model assumptions are tested as part of the modelling process. Thus, whilst in the context of our simulation study it was appropriate to carry out automated implementations of the methods, in practice a considerable amount of time may be required exploring the data, residuals from initial model fits, and fitting further models on the basis of such investigations (in this case, for example, it would rapidly become evident that an exponential covariance structure for the Generalized Least Squares

implementation would be an improvement). In a Bayesian context, this might also involve hands-on tuning or convergence checking as is usual in a single analysis.

Our analysis took a completely heuristic approach to identifying the best methods for spatial regression analysis. The reasons why the various methods performed differently are ultimately a function of the particular mathematical models that underlie the different methods and in some cases are also the result of decisions we had to take about how to implement these methods. For example, Generalized Least Squares is likely to be precise and accurate so long as the spatial covariance matrix can be accurately modelled, but this can be difficult when, for example, the scale of autocorrelation is large and the spatial domain small. We chose to implement Generalized Least Squares in two ways, firstly as Generalized Least Squares-S using the spherical model for the error term, which miss-specifies the error structure, and secondly using a one-parameter version (Simultaneous Autoregressive Models) that has an implicit exponential structure. The resulting misspecification and uncertainty in parameter estimates, for Generalized Least Squares-S explains the inflated Type I error rates given by some Generalized Least Squares-S models in the presence of large-scale spatial autocorrelation: in some realizations of Scenarios 1 and 3 the actual scale of autocorrelation was larger than the spatial domain and consequently the covariance estimate was inaccurate. However, the loss of performance of Generalized Least Squares-S in terms of significance testing was not matched by a corresponding loss of performance in terms of absolute bias and RMSE. Note also that similarly, Generalized Additive Models and GLMM are not single methods as such, because many different types of smoother are available as well as the choice of functional form assumed for the spatial correlation term in Generalized Additive Mixed Models (Hastie & Tibshirani 1990; Wood 2006). Methods like Simple Autoregressive and Spatial Filters emphasize local over global patterns and necessarily perform poorly when assessed against their ability to identify global relationships: indeed Augustin *et al.* (1996) in their original formulation advised against using this method for inference, a warning that has frequently been ignored since [although in certain circumstances it may be local patterns that are ecologically interesting and, if so, these methods and geographically weighted regression may be appropriate (Brunsdon *et al.* 1996; Betts *et al.* 2009)]. In our implementation of the Bayesian method (BCA), we had to establish a process for model selection which involved comparing fits of models with all combinations of covariates. This led to different model selection properties compared with other methods, with lower Type I errors than the nominal 5% rates set elsewhere. For BCA, the low Type I error rates were not associated with high Type II errors, whereas in additional

unreported runs of other methods, setting the nominal significance levels to 1% reduced the Type I error rates at the cost of a considerable increase in the Type II error rates.

The computational cost of unnecessarily fitting a complex model that includes spatial effects is negligible compared with the dangers of ignoring potentially important autocorrelations in the errors and, because spatial and non-spatial methods are equivalent in the absence of spatial autocorrelation in the errors, the precautionary principle suggests models which incorporate spatial autocorrelation should be fitted by default. One possible exception is when the covariance structure of a general Generalized Least Squares or Generalized Additive Mixed Models model is poorly estimated, but this should be apparent from either standard residual inspection or a study of the estimated correlation function. Whilst likelihood ratio tests or AIC could be used to assess the strength of evidence in the data for the particular correlated error models, although these may have low power with small numbers of observations. Note that, strictly, these comments are most relevant when the correlation model has several parameters. Other types of spatial analysis, such as the Generalized Additive Models and Simultaneous Autoregressive Models methods, use a one-parameter 'local smoothing' or 'spatial neighbourhood' approach and it is interesting to observe these models, with a single parameter for the autocorrelation of the errors, performing well compared with our implementation of Generalized Least Squares-S with miss-specified error structure.

Throughout this paper, we have focussed on the analysis of data with normally distributed errors. This deliberate choice reflects the fact that, despite the additional issues introduced by analysis of response variables from other distributions (e.g. presence/absence), all the issues described here are relevant whatever the error distribution. While Bayesian methods using Markov chain Monte Carlo (MCMC) simulations potentially offer a long-term solution to some of these additional issues, their complexities are beyond the scope of this review and instead we refer interested readers to Diggle & Ribeiro (2007). In the meantime, the use of spatial methods that model spatial autocorrelation among the covariates may be more immediately tractable, but equivalent comparative analyses to those presented here would be useful as new methods are developed to analyse spatial data with non-normal distributions.

CONCLUSIONS

In the process of assessing the performance of spatial regression methods, we have shown that some common perceptions are mistaken. We showed that a successfully fitted spatially explicit model may well have autocorrelated residuals (Fig. 5): this does not necessarily indicate a

problem. Similarly, we saw that it is unwise to carry out detrending on dependent variables alone before analysis (Fig. 6). Our work represents an advance in comparative analysis of methods for fitting regression models to spatial data and so provides a more reliable evidence base to guide choice of method than previously available. Whilst our investigations have been far from exhaustive, we hope they represent a step towards repeatable, simulation-based comparisons of methods based on their properties when used to model particular classes of data sets. We have also answered a common question asked by ecologists: parameter estimates from spatially explicit regression methods are usually smaller than those from non-spatial regression methods because the latter are less precise (Fig. 1). Bearing in mind the qualifications and caveats made earlier, we summarize our empirical results by the statements:

- (1) Due to their relatively low precision and high Type I error rates, non-spatial regression methods – notably Ordinary Least Squares regression – should not be used on spatial data sets.
- (2) Removing large-scale spatial trends in dependent variables before analysis (detrending) is not recommended.
- (3) Generalized Least Squares-S, Simultaneous Autoregressive Models, Generalized Additive Mixed Models and Bayesian Conditional Autoregressive performed well in terms of absolute bias and RMSE in the absence of model selection.
- (4) Whilst Type II error rates were generally not a problem for these better performing methods, Generalized Additive Mixed Models and Simultaneous Autoregressive Models have Type I error rates closest to the nominal 5% levels (model selection for BCA is implemented differently and was usually well below 5%).
- (5) After model selection, the performance of Generalized Least Squares-S, Simultaneous Autoregressive Models, Generalized Additive Mixed Models and Bayesian Conditional Autoregressive all improved markedly, with BCA performed best of all in terms of RMSE.
- (6) When fitting large, computationally expensive spatial regression models data partitioning can be effectively utilized with little apparent loss of precision, although one should still be cautious.

ACKNOWLEDGEMENTS

We thank P. Goddard, W. D. Kissling and two anonymous referees for comments on earlier drafts of this manuscript, and M. Schlather for discussion of methods to generate non-stationary random fields. CMB, JJJ, DME & MJB were funded by the Scottish Government.

REFERENCES

- Aitken, A.C. (1935). On least squares and linear combination of observations. *Proc. R. Soc. Edinb.*, 55, 42–48.
- Alpargu, G. & Dutilleul, P. (2003). To be or not to be valid in testing the significance of the slope in simple quantitative linear models with autocorrelated errors. *J. Stat. Comput. Simul.*, 73, 165–180.
- Augustin, N.H., Muggleston, M.A. & Buckland, S.T. (1996). An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.*, 33, 339–347.
- Austin, M. (2007). Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol. Modell.*, 200, 1–19.
- Ayinde, K. (2007). A comparative study of the performances of the OLS and some GLS estimators when stochastic regressors are both collinear and correlated with error terms. *J. Math. Stat.*, 3, 196–200.
- Barry, S. & Elith, J. (2006). Error and uncertainty in habitat models. *J. Appl. Ecol.*, 43, 413–423.
- Beale, C.M., Lennon, J.J., Elston, D.A., Brewer, M.J. & Yearsley, J.M. (2007). Red herrings remain in geographical ecology: a reply to Hawkins *et al.* (2007). *Ecography*, 30, 845–847.
- Beguiria, S. & Pueyo, Y. (2009). comparison of simultaneous autoregressive and generalized least squares models for dealing with spatial autocorrelation. *Glob. Ecol. Biogeogr.*, 18, 273–279.
- Bellier, E., Monestiez, P., Durbec, J.P. & Candau, J.N. (2007). Identifying spatial relationships in the analysis of species distribution coordinates of neighbour matrices (PCNM) and geostatistical approaches. *Ecography*, 30, 385–399.
- Besag, J., York, J.C. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Stat. Math.*, 43, 1–59.
- Betts, M.G., Ganio, L.M., Huso, M.M.P., Som, N.A., Huettmann, F., Bowman, J. *et al.* (2009). Comment on “Methods to account for spatial autocorrelation in the analysis of species distribution data: a review”. *Ecography*, 32, 374–378.
- Bini, L.M., Diniz, J.A.F., Rangel, T.F.L.V., Akre, T.S.B., Albaladejo, R.G., Albuquerque, F.S. *et al.* (2009). Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography*, 32, 193–204.
- Brunsdon, C., Fotheringham, A.S. & Charlton, M.E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr. Anal.*, 28, 281–298.
- Burnham, K.P. & Anderson, D. (2003). *Model Selection and Multi-Model Inference*. Springer, New York.
- Carl, G. & Kuhn, I. (2008). Analyzing spatial ecological data using linear regression and wavelet analysis. *Stoch. Env. Res. Risk A.*, 22, 315–324.
- Carter, S.P., Delahay, R.J., Smith, G.C., Macdonald, D.W., Riordan, P., Etherington, T.R. *et al.* (2007). Culling-induced social perturbation in Eurasian badgers *Meles meles* and the management of TB in cattle: an analysis of a critical problem in applied ecology. *Proc. R. Soc. Lond. B.*, 274, 2769–2777.
- Cheal, A.J., Delean, S., Sweatman, H. & Thompson, A.A. (2007). Spatial synchrony in coral reef fish populations and the influence of climate. *Ecology*, 88, 158–169.
- Cliff, A.D. & Ord, J.K. (1981). *Spatial Processes: Models and Applications*. Pion, London.
- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Chichester, UK.
- Curriero, F.C., Hohn, M.E., Liebholt, A.M. & Lele, S.R. (2002). A statistical evaluation of non-ergodic variogram estimators. *Environ. Ecol. Stat.*, 9, 89–110.
- Dale, M.R.T. & Fortin, M.J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, 9, 162–167.
- Dale, M.R.T. & Fortin, M.J. (2009). Spatial autocorrelation and statistical tests: some solutions. *J. Agric. Biol. Environ. Stat.*, 14, 188–206.
- Davies, R.G., Orme, C.D.L., Storch, D., Olson, V.A., Thomas, G.H., Ross, S.G. *et al.* (2007). Topography, energy and the global distribution of bird species richness. *Proc. R. Soc. Lond. B Biol. Sci.*, 274, 1189–1197.
- Diggle, P.J. & Ribeiro, P.J. (2007). *Model-Based Geostatistics*. Springer, New York, USA.
- Diniz-Filho, J.A.F., Hawkins, B.A., Bini, L.M., De Marco, P. & Blackburn, T.M. (2007). Are spatial regression methods a panacea or a Pandora's box? A reply to Beale *et al.* (2007). *Ecography*, 30, 848–851.
- Dormann, C.F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Glob. Ecol. Biogeogr.*, 16, 129–138.
- Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G. *et al.* (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609–628.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A. *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151.
- Fortin, M.-J. & Dale, M. (2005). *Spatial Analysis*. Cambridge University Press, Cambridge, UK.
- Gaston, K.J. (2000). Global patterns in biodiversity. *Nature*, 405, 220–227.
- Gaston, K.J., Davies, R.G., Orme, C.D.L., Olson, V.A., Thomas, G.H., Ding, T.-S. *et al.* (2007). Spatial turnover in the global avifauna. *Proc. R. Soc. Lond. B Biol. Sci.*, 274, 1567–1574.
- Grand, J., Cummings, M.P., Rebelo, T.G., Ricketts, T.H. & Neel, M.C. (2007). Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecol. Lett.*, 10, 364–374.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, 8, 993–1009.
- Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge, UK.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press, Cambridge, UK.
- Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC, New York.
- Hawkins, B.A., Diniz, J.A.F., Bini, L.M., De Marco, P. & Blackburn, T.M. (2007). Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography*, 30, 375–384.
- Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.*, 54, 187–211.
- Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. *et al.* (2008). Global trends in emerging infectious diseases. *Nature*, 451, 990–993.
- Karagatzides, J.D., Manson, H.R. & Tsuji, L.J.S. (2003). Spatial distribution and performance of *Scirpus americanus* ramets across a

- temperate intertidal marsh resource gradient. *Plant Ecol.*, 169, 215–226.
- Kent, M., Moyeed, R.A., Reid, C.L., Pakeman, R. & Weaver, R. (2006). Geostatistics, spatial rate of change analysis and boundary detection in plant ecology and biogeography. *Prog. Phys. Geogr.*, 30, 201–231.
- Kissling, W.D. & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob. Ecol. Biogeogr.*, 17, 59–71.
- Koenig, W.D. (1999). Spatial autocorrelation of ecological phenomena. *Trends Ecol. Evol.*, 14, 22–26.
- Koenig, W.D. & Knops, J.M.H. (1998). Testing for spatial autocorrelation in ecological studies. *Ecography*, 21, 423–429.
- Kremen, C., Cameron, A., Moilanen, A., Phillips, S.J., Thomas, C.D., Beentje, H. *et al.* (2008). Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, 320, 222–226.
- Kühn, I. (2007). Incorporating spatial autocorrelation may invert observed patterns. *Divers. Distrib.*, 13, 66–69.
- 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., Shiner, S.A. & Franzreb, K.E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monogr.*, 72, 445–463.
- Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D.J. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.*, 10, 325–327.
- Lunn, D.J., Whittaker, J.C. & Best, N. (2006). A Bayesian toolkit for genetic association studies. *Genet. Epidemiol.*, 30, 231–247.
- McMahon, S.M. & Diez, J.M. (2007). Scales of association: hierarchical linear models and the measurement of ecological systems. *Ecol. Lett.*, 10, 437–452.
- Miao, S., Carstenn, S. & Nungesser, M.K. (eds) (2009). *Real World Ecology: Large-Scale and Long-Term Case Studies and Method*. Springer, New York.
- Miller, J., Franklin, J. & Aspinall, R. (2007). Incorporating spatial dependence in predictive vegetation models. *Ecol. Model.*, 202, 225–242.
- Morgan, B.J.T. (1984). *Elements of Simulation*. Chapman & Hall, London.
- Nogues-Bravo, D., Araujo, M.B., Romdal, T. & Rahbek, C. (2008). Scale effects and human impact on the elevational species richness gradients. *Nature*, 453, 216–219.
- Pinheiro, J.C. & Bates, D.M. (2000). *Mixed Effects Models in S and S-Plus*. Springer, New York.
- Pressey, R.L., Cabeza, M., Watts, M.E., Cowling, R.M. & Wilson, K.A. (2007). Conservation planning in a changing world. *Trends Ecol. Evol.*, 22, 583–592.
- R development core team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Saether, B.E., Engen, S., Grotan, V., Fiedler, W., Matthysen, E., Visser, M.E. *et al.* (2007). The extended Moran effect and large-scale synchronous fluctuations in the size of great tit and blue tit populations. *J. Anim. Ecol.*, 76, 315–325.
- Segurado, P., Araujo, M.B. & Kunin, W.E. (2006). Consequences of spatial autocorrelation for niche-based models. *J. Appl. Ecol.*, 43, 433–444.
- Spiegelhalter, D.J., Thomas, A. & Best, N.G. (2000). *WinBUGS Version 1.3 User Manual*. Medical Research Council Biostatistics Unit, Cambridge.
- Vandermeer, J., Perfecto, I. & Philpott, S.M. (2008). Clusters of ant colonies and robust criticality in a tropical agroecosystem. *Nature*, 451, 457–459.
- Velando, A. & Freire, J. (2001). How general is the central-periphery distribution among seabird colonies? Nest spatial pattern in the European Shag *Condor*, 103, 544–554.
- Whittingham, M.J., Stephens, P.A., Bradbury, R.B. & Freckleton, R.P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *J. Anim. Ecol.*, 75, 1182–1189.
- Wiegand, T., Gunatilleke, C.V.S., Gunatilleke, I.A.U.N. & Huth, A. (2007). How individual species structure diversity in tropical forests. *Proc. Natl Acad. Sci. USA*, 104, 19029–19033.
- Wood, S.N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62, 1025–1036.
- Woodroffe, R., Donnelly, C.A., Cox, D.R., Bourne, F.J., Cheeseman, C.L., Delahay, R.J. *et al.* (2006). Effects of culling on badger *Meles meles* spatial organization: implications for the control of bovine tuberculosis. *J. Appl. Ecol.*, 43, 1–10.
- Zhang, L.J., Gove, J.H. & Heath, L.S. (2005). Spatial residual analysis of six modeling techniques. *Ecol. Model.*, 186, 154–177.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Examples of correlated variables with difference spatial scales generated using the custom-built correlated pattern function.

Figure S2 The generation process for a variable with varying scale (non-stationary) autocorrelation.

Figure S3 GLS regression coefficients for univariate and multivariate models with pairs of cross-correlated covariates where the expected regression coefficient for one covariate is 0.5.

Figure S4 Scenario 1: Cross-correlations exist between two pairs of covariates (a) with (b) and (d) with (e).

Figure S5 Scenario 2: Cross-correlations exist between two pairs of covariates (a) with (b) and (d) with (e).

Figure S6 Scenario 3: Cross-correlations exist between two pairs of covariates (a) with (b) and (d) with (e).

Figure S7 Scenario 4: Cross-correlations exist between two pairs of covariates (a) with (b) and (d) with (e).

Figure S8 Scenario 5: Cross-correlations exist between covariates (a) and (b).

Figure S9 Scenario 6: Cross-correlations exist between covariates (a) and (b).

Figure S10 Scenario 7: Cross-correlations exist between covariates (a) and (b).

Figure S11 Scenario 8: Cross-correlations exist between covariates (a) and (b).

Figure S12 Computation time for models of 100 simulations where spatial covariance is either modelled across the whole dataset ('Full'), across (but not between) two even subsets of the data ('Even Split') or two subsets of the data where subset membership was randomly allocated across the spatial domain ('Random split').

Figure S13 The effect of splitting a dataset on precision of parameter estimates for GLS models.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such

materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Editor, James Grace

Manuscript received 10 August 2009

First decision made 5 September 2009

Manuscript accepted 12 November 2009