Author/s:
Crase, B;Liedloff, A;Vesk, PA;Fukuda, Y;Wintle, BA

Title:
Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts

Date:
2014-01-01

Citation:
Crase, B., Liedloff, A., Vesk, P. A., Fukuda, Y. & Wintle, B. A. (2014). Incorporating spatial autocorrelation into species distribution models alters forecasts of climate-mediated range shifts. Global Change Biology, 20 (8), pp.2566-2579. https://doi.org/10.1111/gcb.12598.

Persistent Link:
https://hdl.handle.net/11343/217048

Incorporating spatial autocorrelation into species distribution models alters forecasts of climate mediated range shifts.

Running head: Spatial autocorrelation alters forecasts

BETH CRASE[1*], ADAM LIEDLOFF[2], PETER A. VESK[1], YUSUKE FUKUDA[3] and BRENDAN A. WINTLE[1]

[1] *The Centre of Excellence for Environmental Decisions, School of Botany, University of Melbourne, Parkville 3010, Victoria, Australia*

[2] *CSIRO Ecosystem Sciences, Berrimah 0828, Northern Territory, Australia*

[3] *Department of Land Resource Management, Berrimah 0828, Northern Territory Government, Australia*

* *Corresponding author: Email address: Beth.Crase@unimelb.edu.au*

Telephone: 61 3 83445049

Fax 61 3 93475460

Keywords: spatial autocorrelation, mangrove, sea level rise, climate change, niche model, species distribution model (SDM), Boosted Regression Tree (BRT) models, Generalized Boosting Models (GBM)

Paper type: Primary research article

Abstract

Species distribution models (SDMs) are widely used to forecast changes in the spatial distributions of species and communities in response to climate change. However, spatial autocorrelation (SA) is rarely accounted for in these models, despite its ubiquity in broad-scale ecological data. While spatial autocorrelation in model residuals is known to result in biased parameter estimates and the inflation of Type I errors, the influence of unmodeled SA on species' range forecasts is poorly understood. Here we quantify how accounting for SA in SDMs influences the magnitude of range shift forecasts produced by SDMs for multiple climate change scenarios. SDMs were fitted to simulated data with a known autocorrelation structure, and to field observations of three mangrove communities from northern Australia displaying strong spatial autocorrelation. Three modelling approaches were implemented: environment-only models (most frequently applied in species' range forecasts), and two approaches that incorporate SA; autologistic models and residuals autocovariate (RAC) models. Differences in forecasts among modelling approaches and climate scenarios were quantified. While all model predictions at the current time closely matched that of the actual current distribution of the mangrove communities, under the climate change scenarios environment-only models forecast substantially greater range shifts than models incorporating SA. Furthermore, the magnitude of these differences intensified with increasing increments of climate change across the scenarios. When models do not account for SA, forecasts of species' range shifts indicate more extreme impacts of climate change, compared to models that explicitly account for SA. Therefore, where biological or population processes induce substantial autocorrelation in the distribution of organisms, and this is not modelled, model predictions will be inaccurate. These results have global importance for conservation efforts as inaccurate forecasts lead to ineffective prioritization of conservation activities and potentially to avoidable species extinctions.

Introduction

Species distribution models (SDMs) are used widely to forecast shifts in the spatial distribution of suitable climates for species and communities, under projected changes in climate (Peterson *et al.,* 2002; Thomas *et al.,* 2004; Rapacciuolo *et al.,* 2012; Rojas-Soto *et al.,* 2012). These forecasts are used to inform policy and to prioritize the allocation of resources for conservation (Guisan & Zimmermann, 2000; Kremen *et al.,* 2008; Leathwick *et al.*, 2008). However, if critical assumptions of statistical models are violated, errors of inference and predictive accuracy in modelling the spatial distribution of the focal organism can occur (Legendre 1993; Dormann 2007), leading to poor decision making with a possible outcome of species extinctions and loss of ecosystem services. One such assumption is violated by the presence of spatial autocorrelation in model residuals (Legendre & Fortin, 1989), which can result in errors of ecological inference (Lennon, 2000), reduced SDM predictive accuracy (Betts *et al.,* 2006; Wintle & Bardos, 2006; Dormann, 2007), and poor decisions (Wintle *et al.*, 2011).

Despite the potential for serious errors in inference and prediction arising from the failure to consider SA in model residuals, it is rarely accounted for in models used for forecasting (or hindcasting) climate-mediated species' distributional change (Brown *et al.*, 2011). From a randomly selected set of 44 papers published in 2012 in which SDMs were used to forecast future, or describe past geographic ranges, only three studies explicitly addressed the spatial autocorrelation present in their data or model residuals, and two that tested model residuals for SA found it to be negligible (Chang *et al.,* 2012; Hongoh *et al.*, 2012; Revermann *et al.,* 2012; Tarkhnishvili *et al.*, 2012; de Oliveira *et al.*, 2012) (see Appendix S1 for the articles surveyed).

Spatial autocorrelation in the distribution of organisms can be attributed to extrinsic and intrinsic sources. Extrinsic sources of SA include autocorrelation present in the underlying environmental variables themselves (Legendre, 1993; Lichstein *et al.*, 2002). For example, rainfall, temperature, relative humidity, soil type and depth are all positively spatially autocorrelated. Other sources of extrinsic SA include competitive interactions and interactions with organisms of higher or lower trophic levels (Legendre, 1993) and the spatial configuration of the habitat. In contrast, intrinsic SA arises from population processes and characteristics of the focal species itself and can cause contagion effects (positive autocorrelation) or dispersion effects (negative autocorrelation) (Wintle & Bardos, 2006). Unless SA is specifically addressed within SDMs, only the extrinsic SA associated with the independent environmental variables is captured and represented in the model forecasts, and even then, only if all relevant spatially structured environmental variables are included in the model.

Several methods to account for SA within the model structure have been developed (Keitt *et al.*, 2002; Dormann *et al.*, 2007). However, only a subset of these can be applied within all of the commonly utilized statistical modelling approaches used to develop SDMs, including generalized linear models (GLM, McCullagh & Nelder, 1983), generalized additive models (GAM, Hastie & Tibshirani, 1987), and newer machine learning techniques such as boosted regression trees (BRT, Breiman *et al.,* 1984; De'ath & Fabricius, 2000), and maximum entropy methods (MaxEnt, Phillips *et al.,* 2004). For example, spatial eigenvector mapping (SEVM) cannot be implemented across this range of modelling approaches (see Dormann *et al.*, 2007). In contrast, the autologistic approach (Besag, 1974; Augustin *et al.*, 1996) and the residuals autocovariate (RAC) approach (Crase *et al.*, 2012) provide relatively simple ways to account for SA within many of these modelling approaches, via the calculation of an autocovariate term that is fitted to the SDMs along with the environmental explanatory variables.

The implications of including or ignoring SA when forecasting future species distributions have not been widely explored (Swanson *et al.*, 2013; Record *et al.*, 2013). However, quantifying the difference between forecasts that account for SA and contrasting them with predictions from models that fit only environmental explanatory variables is important because these environment-only models have been used to forecast range shifts for thousands of organisms. The sensitivity of model forecasts to assumptions about SA would call into question the accuracy of these forecasts and their utility in conservation planning. Therefore, the aims of the present paper are: (i) to quantify the magnitude and direction of difference between forecasts provided by the usual implementation of SDMs and those that account for SA; (ii) explore the circumstances under which large differences in prediction between environment-only and SA models are likely to occur, (iii) discuss the implications for conservation modeling of including or ignoring SA in distribution forecasts, and (iv) to offer some guidance on modelling approaches to consider when model residuals are strongly spatially autocorrelated. To address these aims we compare distribution change forecasts from environment-only and spatial autocorrelation models fitted to simulated datasets and mangrove community distribution data from northern Australia.

Methods

Two Bernoulli distributed (presence/absence) datasets, one simulated and the other observed, were used to quantify differences among model forecasts. Models were fitted to the initial distribution of focal organisms and used to forecast the distribution of the organisms under projected future environmental conditions.

Simulated dataset

Dormann *et al.* (2007) simulated ten presence/absence datasets across a grid of 1108 cells for a hypothetical species ("snouter") with a known, spatially correlated error structure and a linear response to an explanatory variable called 'rain'. A second, uninformative explanatory variable, “djungle”, was included as a candidate explanatory variable by Dormann *et al.* (2007) in order to test type I error rate in the modelling methods they evaluated. Including both known informative and uninformative independent variables (rain and djungle) in our analyses enabled us to test how well the modelling approaches dealt with the presence of a noise variable.

Mangrove forest dataset

A dataset was derived from the mapped distribution of three mangrove communities occurring in Woods Inlet, a 9 x 9 km area within a large (20, 400 ha), relatively undisturbed mangrove forest adjacent to Darwin Harbor in northern Australia. This extensive mangrove forest has experienced triple the global mean rate of recorded sea level rise over the past 18 years (8.3 mm/yr, 1992 - 2010, Australian Baseline Sea Level Monitoring 2011). Darwin Harbor has semi-diurnal macro tides with tidal amplitudes to 8 m (Williams *et al.,* 2006). The maximum daily temperatures range from 30.5° to 33.3° C, and the average annual rainfall of 1728 mm falls during the wet season from November to March (Australian Bureau of Meteorology, 2012).

The study area within Wood’s Inlet was divided into 25 x 25 m grid cells, and each cell was assigned the presence or absence of the mangrove communities based on a map of the communities produced by Brocklehurst and Edmeades (1996). This represents complete (or saturated) sampling where a value is provided for every cell within the study area. The three mangrove communities are each dominated by a single tree species, and account for 82% of the Darwin Harbor mangrove forest (Brocklehurst & Edmeades, 1996). The community dominated by *Sonneratia alba* Sm. covers 5% of the mangrove forest and is restricted to locations with the lowest elevation. The community dominated by *Rhizophora stylosa* Griff occupies 33% of the forest and occurs in locations with an intermediate elevation, while the *Ceriops tagal* Perr. dominated community covers approximately 44% of the forest and occurs mainly in areas with the highest elevation (Brocklehurst & Edmeades, 1996).

Three explanatory variables were fitted to the models: the salinity of inundating water in the wet season, slope of the terrain and the hydroperiod, which is the duration of inundation expressed as a percent of time. These variables were selected as previous work has shown a significant relationship between these variables and mangrove community distribution (Crase *et al.*, 2013). Pearson correlations between the variables were calculated before fitting the models to ensure that none of the variables were highly correlated (R <0.3, Appendix S2). The salinity of the inundating water during the wet season, was derived for each grid cell in the study area from an interpolation of 689 measurements of water salinity made throughout the harbor during the wet season of 2009 and 2010 (unpublished data, Aquatic Health, Northern Territory Government). Water salinity was measured at the peak of the wet season in both years, and recorded with two hydrolabs (Datasonde4a and DS5X, Hydrolab, Colorado, USA) 0.2 m below the water surface in parts per thousand (ppt). Wet season water salinity ranged from 4 to 31 parts per thousand (ppt).

For each grid cell in the study area we calculated the slope in degrees based on a DEM (digital elevation model) derived from LiDAR (Light detection and ranging) with a vertical resolution of 1 cm ± 19 cm and a grain size of 1 m (Temby, 2009). Slope varied from 0 to 30 degrees.

Hydroperiod was calculated for each grid cell in the study area by grouping tide heights in Darwin Harbor recorded every 20 minutes for a year, into 1 cm elevation classes (D. Williams, unpublished data), then matching the elevation of the grid cells to tide height and extracting the percent of time inundated. The hydroperiod ranged from 0% (never inundated) to 100% (continuously inundated). All variables were manipulated and derived using ArcGIS (ESRI vers.10.1).

Simulating environmental change scenarios

For the simulated dataset, four scenarios of climate change-mediated rainfall increase were produced by increasing the explanatory variable, ‘rain’, for each cell in the study area, by steps of 100 mm, until rain was 600 mm greater than in the baseline simulation. The second explanatory variable, djungle, remained fixed. For the mangrove communities, two simulated scenarios were produced by increasing the sea level by 0.5 and 1 m, and recalculating the hydroperiod for each cell in the study area. The other explanatory variables (slope and wet season water salinity) were held at current values. A caveat of the study should be noted here, that some extrapolation beyond the empirically observed relationships between the response variable and rain or inundation (for the simulated and observed datasets, respectively) is unavoidable. For the simulated dataset, the shift upward is approximately 50% of the observed range of rainfall and while the lower half of the rainfall range when increased by the maximum (6 by 100mL steps) is still within the observed range, most sites above the midpoint (approximately 650mL) will exceed the range. This can be a source of model uncertainty, which propagates through to the predictions. However, the three modelling approaches are compared using the same assumptions about the relationship between the response and explanatory variables, and this extrapolation is unlikely to bias one method over the others.

Modelling approaches

We implemented Boosted Regression Tree models (BRT), which are a class of machine learning statistical models able to automatically capture interactions between variables and deal with complex non-linear relationships (De'ath & Fabricius, 2000; Elith *et al.,* 2008). Each tree was fitted with three nodes and a bag fraction of 0.5 using the gbm-package library (Ridgeway, 2006) and code from Elith *et al.* (2008), in R (vers 2.11.1; R Core Development Team, 2012). A minimum of 1000 trees was fitted to each model. For detail on the BRT model fitting procedures refer to Elith *et al.* (2008).

Three classes of models were implemented. First, environment-only models that do not explicitly account for SA; second, RAC models that utilize an autocovariate derived from the residuals of an environment-only model (Crase *et al.,* 2012), and thirdly, for the sake of comparison with a widely applied SA modelling approach, we fitted standard autologistic models that included the environmental explanatory variables and an autocovariate representing SA in the response variable (Besag, 1974; Augustin *et al.,* 1996). The same environmental explanatory variables were fitted to all models, so the difference between these three classes of models was whether or not an autocovariate term was included, and the way the term was derived.

The RAC and autologistic models both derive a term reflecting the spatial autocorrelation present, however, for the autologistic method this term represents SA in the observation data (Augustin *et al.*, 1996), while for the RAC method this term represents SA in the model residuals (Crase *et al.*, 2012). The autocovariate terms were calculated for a 3 x 3 cell neighborhood surrounding each cell in the study area. The autologistic autocovariate term was calculated such that,

$$autocovariate_i = \frac{1}{|\boldsymbol{N}_i|}\sum_{j\in \boldsymbol{N}_i} y_j \qquad \text{Equation 1}$$

where y_j is the value of the response variable at location j, where j is a cell within the set $\boldsymbol{N}_i$ that forms neighborhood $\boldsymbol{N}$ for location i. For each location i, a neighborhood is defined, so the summation is across the set of cells in the neighborhood and divided by the number of cells in the neighborhood. The autocovariate and the environmental explanatory variables were then fitted to the BRT models at the same time as the explanatory variables.

The RAC autocovariate term was calculated such that,

$$autocovariate_i = \frac{1}{|\boldsymbol{N}_i|} \sum_{j \in \boldsymbol{N}_i} (y_j - q_j) \qquad \text{Equation 2}$$

where q_j is the estimated probability of occurrence at site j derived from an environment-only model, y_j is the value of the response variable at location j, summed across all cells within the defined neighborhood $\boldsymbol{N}_i$. The RAC autocovariate term and the environmental explanatory variables were then fitted to the BRT models.

For all fitted models, predictive performance was assessed using two cross-validated performance metrics, AUC (Area Under Curve of the Receiver Operating Characteristic, ROC) and the percent of deviance explained. AUC indicates the ability of a model to correctly distinguish between pairs of occupied and unoccupied sites, and ranges from 0.5 (random) to 1 (perfect discrimination), with scores above 0.7 considered to indicate useful models (Swets, 1988; Pearce & Ferrier, 2000). The percent of deviance explained by the model shows the goodness of fit between modelled and observed values, and reflects the match between the actual and predicted frequency of occurrence of the organism (Ferrier & Watson, 1997).

To ensure that the data used to fit the models was not also used to test model performance we calculated the metrics of model performance using ten-fold cross-validation (Stone, 1974).

The full dataset was divided into ten portions, with one portion held-out for testing model predictions and the other nine portions combined and used to fit the models. Performance metrics were calculated by comparing the model predictions to the observation data in the tenth held out portion. We repeated this process until the models had been fitted and tested ten times to calculate the cross-validated AUC and deviance reduction statistics as averages (and standard deviations) across the 10 tests. Cross-validation does not provide a true out-of-sample test of the predictive performance of the model, but rather an estimate of out-of-sample predictive performance (Stone 1974).

Moran's Index was calculated to quantify the SA remaining in model residuals of each of the models fitted for the Snouter and the mangrove communities. Moran's Index ranges from 1 (strong positive SA) to -1 (strong negative SA), with zero indicating a random pattern with no SA (Cliff & Ord, 1981). Here, Moran's Index was calculated across a range of distances and presented as correlograms, based on 1000 cells sampled from the 17250 cells in the study area; and for a neighbourhood size of 3 x 3 cells, with the neighbourhood size selected to present the greatest residual values in SA in order to illustrate the reduction in SA when the autocovariate methods are applied (Crase *et al.*, 2012).

Producing and comparing forecasts under environmental change

The model of Snouter was fitted to data for the 'current' rainfall scenario. This model was then used to generate a probability of occurrence for each cell in the study area based on current rainfall, and for the future rainfall scenarios. Likewise, the mangrove community models were fitted to the current values of the explanatory variables, and then used to generate a probability of occurrence of each mangrove community at current sea level, and for a sea level rise of 0.5 m and 1 m. To derive $E_{m,}$, which is the expected number of cells occupied by the focal organism (Snouter or each mangrove community), for one of the climate scenarios *m*, the probability of occurrence was summed across all cells in the study area,

$$E_m = \sum_{i=1}^{n}(p_{i,m}) \qquad \text{Equation 3}$$

where *n* is the number of cells in the study area, $p_{i,m}$ is the probability of occurrence of the focal organism for cell *i*. This threshold-free approach to calculating the expected number of cells occupied avoids the problems associated with selecting a threshold to convert each probability of occurrence to a presence or absence (see Liu *et al.*, 2005; Jiménez-Valverde & Lobo, 2007). To determine the variability in the estimates of E_m, we bootstrapped the full dataset - resampling 1000 times, then fitted the three modelling approaches to each reampled dataset, and produced forecasts under each of the change scenarios. This enabled us to calculate mean ± standard error of E_m.

Results

Spatial autocorrelation in model residuals

All modelling approaches accounted for some of the spatial autocorrelation present in the response variable for both the simulated and observed datasets, although at short distances, substantial SA remained in the residuals of the environment-only models (Figure 1). Moran's Index calculated for a neighborhood size of 3 x 3 cells showed that the autologistic and RAC models controlled SA, reducing it to between -0.01 and -0.09, while substantial SA remained in the residuals of the environment-only model (0.3-0.4, Table 1). While SA was reduced by the autologistic and RAC models some noise remained in the model residuals at all distances (Figure 1).

Model performance

For the simulated data, the RAC and autologistic models had strong model performance, as indicated by AUC scores above 0.95. The environment-only model, however, performed poorly (AUC 0.68) (Table 2a). For the mangrove community data, all modelling approaches had very high AUC scores (Table 2b), demonstrating that all models, including the environment-only model, could accurately distinguish between pairs of occupied and unoccupied sites. A greater proportion of deviance was explained by the RAC and autologistic approaches compared to the environment-only model, for both the simulated and mangrove datasets (Table 2).

Relative variable importance

Relative variable importance (RI) indicates the number of times the variable is selected to define a bifurcation in a BRT model. For the simulated 'snouter' data, the RI of the predictor variable rain ranged from 78.9-93.9% for the env-only model, to 4.3-13.1% for the RAC approach and 4.9-9.1% for the autologistic approach. In contrast, the uninformative variable djungle, was rarely selected by the model fitting algorithm and consequently had low relative importance (env-only 6.1-21.1%, RAC 0.3-2.6%, autologistic 1.7-4.5%). This compares well with Dormann *et al.*, (2007) where ‘djungle’ was a non-significant explanatory variable in six modelling approaches applied to control SA.

For the observed mangrove data, the autocovariate terms had the greatest RI in the RAC and autologistic approaches (range: 70.4-97.7%). The RI of slope and salinity of the inundating water was moderate to low in all approaches (water salinity: 3.3-38.9%; slope 4.9-24.4%). The predictor variable hydroperiod had a high RI in the env-only models (41.1-55.6%), moderate in the RAC approach (6.4-19.8%) and low in the autologistic approach (1.2-1.5%) (Appendix S3). The key point is that a change in the model structure (ie when an autocovariate is included) alters the inference drawn from the model. The apparent importance of the hydroperiod in structuring community distribution is diminished when autocovariate terms are included, particularly so for the autologistic approach and less so for the RAC approach.

The partial plots illustrate how the modelling approach influences the inferred relationship between the probability of occurrence of by particular mangrove community and the predictor variable hydroperiod (Figure 2). The shape of the relationships is similar for the environment-only and RAC approaches for *C. tagal* (Figure 2 a, d), although the relative importance of the hydroperiod is greater in the environment-only model (54.4%) than the RAC model (19.8%). In contrast, for the autologistic model, the relationship is much less pronounced between occurrence and hydroperiod (Figure 2, g) and the relative variable importance is very low (1.5%). Similar patterns were recorded for the *R. stylosa* and *S. alba* communities (Figure 2).

Expected occupancy for scenarios of higher rainfall or higher sea level

For the current level of rainfall, the expected number of cells occupied by the simulated organism was similar for all three modelling approaches (mean ± standard error, E_m : environment-only 843.6 ± 29.4; autologistic 843.7 ± 29.3 RAC 843.4 ± 29.4). As rain increased across the six future scenarios, Snouter was forecast to change in occupancy (Figure 3a). However, the expected number of cells occupied by Snouter differed among the models and the magnitude of the difference in forecasts among the modelling approaches increased with each increment in rainfall. For a 600 ml increase in rain, the environment-only model forecast the greatest decrease in area occupied by Snouter (E_m : 806.7 ± 36.7). In contrast, the autologistic model forecast little decline in area occupied (E_m : 846.4 ± 27.0). The RAC model forecasts were intermediate between those of the environment-only model and the autologistic model (E_m : 824.0 ± 29.5). The magnitude of the difference in forecasts among the modelling approaches increased with each increment in rainfall, and forecasts differed least between the RAC and autologistic approaches, followed by the RAC and environment-only approaches (Figure 3b).

For the mangrove communities, there was little difference between the forecasts of the three modelling methods at the current sea level (Figure 4). All modelling approaches estimated a similar expected number of occupied cells (Figure 5, see also Appendix S4 for values of the mean and standard errors). For example, for the *S. alba* community the mean E_m (± standard error) were 835.0 ± 0.47, 845.11 ± 0.28 and 848.68 ± 0.27, for the environment-only, RAC and autologistic approaches, respectively. However, the expected number of occupied cells forecast over sea level rises of 0.5 and 1 m, diverged among modelling approaches with each increment in sea level. The environment-only model forecast an increase for *S. alba* and *R. stylosa*, but a steep decline for the *C. tagal* community (Figure 5, Supporting information, Appendix S4 and S5). The autologistic approach, in contrast, forecast little change in the expected number of occupied cells for each community under conditions of higher sea level. The RAC model forecasts were intermediate between the environment-only and autologistic forecasts. For example, for the RAC approach, the *S. alba* and *R. stylosa* communities were forecast to increase in area after future sea level rises (Figure 5 a, b), while the *C. tagal* community declined (Figure 5 c), but less so than those of the environment-only model forecasts.

The differences in the magnitude of spatial change in suitable habitat forecast by the environment-only models and SA models was substantial. For all mangrove communities, the environment-only and autologistic forecasts differed the most, and these differences increased with each incremental rise in sea level, irrespective of whether the forecast area occupied increased (the *S. alba* and *R. stylosa*) or decreased (*C. tagal*). Furthermore, the magnitude of difference between the environment-only and RAC forecasts varied for each mangrove community, with the greatest differences predicted for the *R. stylosa* dominated community. For a 1 m rise in sea level, the area occupied by the *S. alba* community forecast by the autologistic and RAC approaches were 74.6 and 84.1% of that forecast by the environment-only model, respectively. The area occupied by the *R. stylosa* community forecast was 57.1% for the autologistic approach and 82.0% for RAC approach compared to the area forecast by the environment-only model. The *C. tagal* community, for which a range contraction was forecast (unlike the other two communities which were forecast to expand in suitable habitat), the suitable area forecast by autologistic and RAC approaches were, respectively, 229.7 and 147.3% of the area forecast by the environment-only model, with this model providing the most extreme forecasts of range contraction for this community. In addition, the spatial distribution of the greatest differences between the forecasts provided by different modelling approaches was concentrated at the leading and trailing edges of the species distribution forecasts (Supporting information, Appendix S6).

Discussion

We found substantial differences in the magnitude of spatial change in suitable habitat forecast by the environment-only models and SA models, despite all models predicting similar ‘current’ distributions, for both the simulated and empirical datasets. The forecasts made by the environment-only models depend entirely on the environmental variables fitted to the models, and therefore, the spatial distribution of the simulated organism and mangrove communities tracked change in the environmental variables. Therefore, increases in rainfall or sea level produced a substantial change in distribution of the organisms. The forecasts from models incorporating SA differed from those of the environment-only models because the SA in the current distribution of the target organism influences future forecasts and usually diminishes the apparent importance of the environmental variables. Furthermore, the forecasts provided by models that do and do not account for SA diverge with each increment in environmental change, and these changes are most apparent at the leading and trailing edges of species' distributions. Previously published forecasts of the spatial distribution of thousands of species have been developed without incorporating SA into these SDMs (Appendix S1). If biological or population processes induce substantial autocorrelation in the distribution of those organisms, and this is not modelled, then these model predictions may be inaccurate. The importance of this inaccuracy depends on the individual species, the environment and decision context.

The forecasts differed between the two modelling approaches we applied to account for SA due to the way the autocovariate was calculated in each method. In the autologistic approach the autocovariate is estimated directly from observations of the organism and therefore emphasizes SA in species distribution, over the influence of the environmental explanatory variables. In contrast, the RAC approach estimates the autocovariate from the residuals of an environment-only model, therefore capturing the influence of the environmental variables before accounting for the remaining SA. This seems a sensible approach because it explicitly targets the assumptions of statistical models, namely that the errors are independent and identically distributed, an assumption violated when SA remains in the model residuals, rather than in the observations of species distribution. In addition, the RAC model fits are smoother and more regularized (Figure 2) than the environment-only approach, which could result in a more realistically representation the relationship between the dependent and independent variables. Therefore, dealing with SA within the model structure can allow more ecologically realistic relationships to be fitted.

The three approaches applied here, the environment-only, autologistic and RAC approaches, represent three hypotheses about the main drivers of the spatial distribution of organisms. For the environment-only approach, the environmental explanatory variables are considered as the sole drivers of the spatial distribution of the organism, and SA present in the observations of the organism are implicitly assumed to arise from SA in the environmental variables themselves. That is, the environment-only approach implicitly assumes that all SA is extrinsic. This is the implicit assumption of 90% of forecasting studies published in 2012 (Appendix 1). The autologistic approach combines intrinsic and extrinsic processes leading to SA in the autocovariate, with the result of de-emphasizing the influence of the environmental variables. This is why the autologistic model forecasts of distribution differ little between the current and future scenarios, despite substantial change in the environmental variables of rain and sea level in the simulated and empirical datasets. In contrast, while the RAC approach prioritizes the effect of environment over the other sources of spatial structuring, the environmental variables are not assumed to be the sole drivers of species distributions. The RAC approach incorporates the effect of other sources of spatial structuring, such as intrinsic biological processes that induce aggregation or dispersion. Consequently, the RAC model forecasts tend to be intermediate between the two extremes formed by the environment-only and autologistic model forecasts.

These three hypotheses are not exhaustive, and an additional hypothesis is provided by models that consist only of interpolations of observations (Bahn & McGill, 2007). This approach does not utilize environmental variables at all and implicitly assumes that the SA in the response variable (i.e. the distribution of the focal species) is the best predictor of species' distribution. Extrinsic and intrinsic SA could be correlated, leading to the spurious detection of a relationship between the distribution of the organism and the environmental variables, thereby incorrectly inflating estimates of the importance of the environmental variables (Bahn & McGill 2007, Lennon 2000, Diniz-Filho *et al.*, 2003).

The explanatory variable hydroperiod was a strong predictor of mangrove community distribution in the environment-only approach and partially in the RAC implementation. This is supported by experimental evidence indicating that the hydroperiod differentially influences the growth rate of mangrove species and these differences in growth rate lead to species, and therefore plant communities, occupying different sections of the gradient in hydroperiod (Elison & Farnsesworth 1997, Kitaya *et al.*, 2002, He *et al.*, 2007, Cardona-Olarte *et al.*, 2006). For Darwin Harbor, the relationship between the probability of occurrence and hydroperiod is unique for each of the mangrove communities in our study, as shown in the partial dependence plots (Figure 2) (see also Crase *et al.* 2013). As the hydroperiod increases under simulated conditions of a higher sea level, these relationships, which are defined in ecological space, translate to a unique response from each mangrove community in geographic space, and therefore to the differences in the expected number of cells occupied (Figure 5). When we assume that the SA in mangrove communities is due to intrinsic sources and we implement the autologistic approach, the modelled relationship between community occurrence and hydroperiod is much diminished compared to env-only models, as shown by a low variable importance score, the attenuated peak in the fitted functions for *C. tagal*, and *R. stylosa*, and flat relationship between the *S. alba* community occurrence and hydroperiod. This difference in SA within each of the communities translates to the differences in forecasts between the mangrove communities.

Species distribution models used for forecasting are usually implemented with either (a) no dispersal or (b) unlimited dispersal, which assumes that organisms can reach all suitable habitat (Araújo & Peterson 2012). However, dispersal is a source of SA likely to critically influence the establishment of species in new locations, and therefore, the accuracy of model projections across time and space in changing environments (Elith & Leathwick 2009). For example, the dispersal capacity of plants has been shown to be positively correlated with SA in environment-only model residuals (Dirnböck & Dullinger 2004), and this would influence the magnitude of change forecast for the species. In addition, simulations have shown that species with low growth rates have a spatial distribution more strongly influenced by dispersal capacity (Bahn *et al.,* 2008). Therefore, species with a limited dispersal capacity and a slow growth rate are likely to be less accurately modelled with the widely implemented environment-only approaches, and SA models should be considered for species with these traits.

Approaches incorporating dispersal, migration and biotic interactions into SDMs are frequently advocated (Araújo & Peterson 2012; Thuiller *et al.* 2008; Elith & Leathwick 2009; Dormann *et al.* 2012), and there has been some progress in this area. Hybrid or integrated approaches to modelling can be used to explicitly incorporate some ecological or biological processes into correlative SDMs (Dormann *et al.* 2012), and some of these processes are sources of intrinsic SA, for example dispersal, patch dynamics and biotic interactions (see for example, Conlisk *et al.*, 2012; Fordham *et al.*, 2012; Meier *et al.* 2012, Pagel & Schurr, 2012; Zurell *et al.*, 2012). While these studies do not explicitly measure nor address SA, it is likely that incorporating such processes into SDMs could lead to reduced SA in model residuals. However, these mechanistic-correlative hybrid approaches to dealing SA are unlikely to be practical for the vast majority of species due to the intensive data required to parameterize such models (Kearney *et al.*, 2010). Therefore, despite their limitations, correlative SDMs will continue to be relied on to produce forecasts of species' ranges and habitat availability and it is critical that the underlying assumptions of SDMs with regard to SA are made explicit so that results and implications of those can be appropriately interpreted.

From a statistical perspective we recommend that the SA in model residuals is at least quantified and reported. SA in residuals does not only potentially indicate intrinsic SA, it can also indicate missing environmental variables (Diniz-Filho *et al.*, 2003; Santika & Hutchinson 2009). Where processes such as dispersal and biotic interactions are known to produce strong SA in the focal organism and model residuals are spatially autocorrelated, at least one of the SA modeling approaches should be implemented in an attempt to capture those biological processes that induce aggregation. Future research directions include exploring the threshold at which SA in model residuals unacceptably influence statistical tests, inference drawn from the models and spatial forecasts. Intrinsic and extrinsic sources of SA may influence this level of unacceptable SA in different, and perhaps, in non-linear ways. Currently, we know that SA in model residuals is a problem, but no guidance can be offered about the intensity of the problem. In some cases, the threshold of unacceptable SA could be determined by the decision context and the level of acceptable uncertainty in forecasts.

A related issue is overfitting the models, which is a general issue within the field of correlative models used in ecology (Araújo & Guisan 2006). Overfitting leads to models that may have poor predictive performance when used for forecasting for different regions or time periods from the one used to fit the model. In modelling approaches that incorporate SA, a critical assumption is that the processes that lead to the observed SA will remain stationary, just as the relationships between species occurrence and the environmental drivers are assumed to remain constant when forecasting across time and space. Where correlations with the biological processes that cause aggregation or dispersion in species occurrence are not stationary, model forecasts are likely to be poor. This would be especially true for SA models that emphasize intrinsic SA. In contrast, the RAC approach may be more appropriate due to its apparent compromise between extrinsic and intrinsic processes.

Other studies comparing the accuracy of environment-only models with SA models are mixed. Swanson *et al.*, (2013) used a historic observational dataset of 99 plant species from the early 20th century to train environment-only and SA models and compared model forecasts to an independent dataset collected between 2000 and 2005. As in the present study, they found that the models accounting for SA produced smaller range shifts than did the environment-only models. They also found that while accounting for SA improved the projection accuracy for some species, it was not universally important. Similarly, Record *et al.*, (2013) developed distributional models for two tree genera using their current distribution and compared hindcasts of distribution with palynological data at 1000 year intervals for 8000 years. They found that models accounting for SA were more accurate for one genus (*Tsuga*), but that the environment-only model was more accurate for the other (*Fagus*). These studies provide evidence that, for some organisms, accounting for SA in the distribution of organisms leads to improvements in model accuracy, but these improvements cannot be generalized across all species nor genera. Accuracy is likely to depend on the degree of SA remaining in model residuals of non-spatial models, the sources of SA, and whether these sources are stationary, that is, whether they remain constant over time.

In the absence of information about intrinsic and extrinsic sources of SA and their stationarity over time, the aims of the forecasting study should inform the selection of environment-only or SA models. For example, if forecasting the greatest possible range shifts for species is the primary goal (i.e. the cost of false negatives is high), the environment-only model is most likely to produce forecasts that characterize the outer limits of possible range changes, compared to the SA model forecasts. Alternatively, models accounting for SA should be used to provide forecasts if the cost of the false positives is greater than false negatives, or if the aim is to provide predictions that take better account of dispersal or other intrinsic limitations to a species capacity to occupy new geographic locations. Approaches exploring the upper and lower bounds of forecasts or ensemble approaches including models accounting for SA could be implemented. The key point is that the decision to include or ignore SA in SDMs must be made, and the current default to ignore it is a decision that may no longer be defensible.

Here we demonstrate that environment-only models, which are widely applied in ecology for forecasting species' distribution, produce forecasts with the greatest range shifts, in comparison to modelling approaches that incorporate SA. The failure to incorporate SA into models used for predictions, not only violates the assumptions of statistical models when SA remains in the model residuals, it may lead to overestimation of climate mediated spatial change in species' distribution or suitable range. A change in the implementation of SDMs to account for SA should lead to predictions of species distributions that are more realistic, and to better prioritization of management activities and resource allocation for conservation.

Acknowledgements

BC was supported by an Australian Postgraduate Award, and the National Environmental Research Program (NERP) Decisions Hub. BW is supported by an Australian Research Council Future Fellowship (FT 100100819) and the ARC Centre of Excellence for Environmental Decisions. We thank Pablo Luis Sordo Martinez and Tak Fung for their assistance in running the bootstrapped models with the High Performance Computing center at the National University of Singapore. Comments from Simon Ferrier, Reid Tingley, Michael Bode and Chris Jones improved earlier versions of the manuscript.

Biosketch

Beth Crase has research interests in spatial ecology, with a focus on tropical ecosystems and forecasting the impacts of climate change on plant species' distributions. Her current research focuses on forecasting distributional change of species in response to climate change and sea level rise; improving the implementation of statistical models used for forecasting and investigating dynamic competitive interactions between species.

Author contributions: B.C., A.L., B.W. and P.V. conceived the ideas; B.C. & Y.F. analyzed the data; and B.C. lead the writing, with all authors contributing.

Supporting Information

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

Appendix S1 Papers on species distribution forecasts surveyed for their incorporation of spatial autocorrelation into the models used for forecasting.

Appendix S2 Values for Pearson correlations (and p-values) between the three explanatory variables (salinity of the inundating water, slope, and duration of inundation) used in the models of mangrove community distribution.

Appendix S3 Relative variable importance (RI) for all variables fitted to environment-only, RAC and autologistic approaches, for the three mangrove communities dominated by *Ceriops tagal*, *Rhizophora stylosa* or *Sonneratia alba*.

Appendix S4 Mean number of cells occupied (E_m) with standard errors and standard deviations, for the mangrove communities for each sea level rise scenario.

Appendix S5 Modelled distribution of the mangrove communities dominated by *Ceriops tagal*, *Rhizophora stylosa* or *Sonneratia alba* in Wood's Inlet, Northern Australia, at current sea level and after a 1 m sea level rise, as provided by environment-only, RAC (residuals autocovariate) and autologistic models.

Appendix S6 Spatial differences in forecasts of the probability of occurrence, between environment-only, RAC and autologisitc approaches, for three mangrove communities dominated by *Ceriops tagal*, *Rhizophora stylosa* or *Sonneratia alba*, after a 1 m rise in sea level.

References

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Araújo MB, Peterson AT (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527-1539.

Augustin NH, Mugglestone MA, Buckland ST (1996) An autologistic model for spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339–347.

Australian Baseline Sea Level Monitoring (2011) *Australian baseline sea level monitoring project file information and instructions*, IDO71014.North-ern Territory Bureau of Meteorology. Available at: http:// www.bom.gov.au/oceanography/projects /abslmp/data/data.shtml (accessed 21/10/12).

Australian Bureau of Meteorology (2012) *Climate statistics for Australian locations*. Australian Bureau of Meteorology. Available at: http://www.bom.gov.au/ climate/averages/tables/cw_014015.shtml (accessed 21/10/12).

Bahn V, McGill BJ (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology & Biogeography*, **16**, 733–742.

Bahn V, Krohn WB, O’Connor, RJ (2008) Dispersal leads to spatial autocorrelation in species distributions: A simulation model. *Ecological Modelling*, **213**, 285–292.

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistics Society, Series B*, **36**, 192–236.

Betts MG, Diamond AW, Forbes GJ, Villard MA, Gunn JS (2006) The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecological Modelling*, **191**, 197–224.

Breiman L, Friedman JH, Olshen RA, Stone CG (1984) *Classification and regression trees*. Wadsworth International Group, Belmont.

Brocklehurst P, Edmeades B (1996) *The mangrove communities of Darwin Harbour*. Department of Lands, Planning and Environment, Resource Capability Assessment Branch. Northern Territory Government, Darwin.

Brown, CJ, Schoeman, DS, Sydeman, WJ, Brander, K, Buckley, LB, Burrows, M, Duarte, CM, Moore, PJ, Pandolfi, JM, Poloczanska, E, Venables, W, Richardson, AJ (2011) Quantitative approaches in climate change ecology. *Global Change Biology*, **17**, 3697-3713.

Cardona-Olarte P, Twilley RR, Krauss KW, Rivera-Monroy V (2006) Responses of neotropical mangrove seedlings grown in monoculture and mixed culture under treatments of hydroperiod and salinity. *Hydrobiologia*, **569**, 325-341.

Chang J, Chen D, Ye XP, Li SH, Liang W, Zhang ZW, Li M (2012) Coupling genetic and species distribution models to examine the response of the Hainan Partridge (*Arborophila ardens*) to late Quaternary climate. *Plos One*, **7**, e50286.

Cliff AD, Ord JK (1981) *Spatial processes: Models and applications.* Pion, London.

Conlisk E, Lawson D, Syphard AD, Franklin J, Flint L, Flint A, Regan HM (2012) The roles of dispersal, fecundity, and predation in the population persistence of an Oak (*Quercus engelmannii*) under global change. *Plos One*, **7**, e36391.

Crase B, Liedloff AC, Wintle BA (2012) A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, **35**, 879–888.

Crase B, Liedloff AC, Vesk PA, Burgman M, Wintle BA (2013) Hydroperiod is the main driver of the spatial pattern of dominance in mangrove communities. *Global Ecology & Biogeography*, **22**, 806-817.

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for the analysis of complex ecological data. *Ecology*, **81**, 3178–3192.

Diniz-Filho JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology & Biogeography*, **12**, 53–64.

Dirnböck T, Dullinger S (2004) Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, **15**, 77–84.

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

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

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

Elith J, Leathwick JR (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology & Systematics*, **40**, 677–697.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

ESRI (2011) *ArcGIS desktop*, release 10.1 Environmental Systems Research Institute, Redlands, CA.

Ferrier S, Watson G (1997) *An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity*. Biodiversity Group, Environment Australia, Sydney.

Fordham DA, Akçakaya HR, Araújo MB, Elith J, Keith DA, Pearson R, Auld TD, Mellin C, Morgan JW, Regan TJ, Tozer M, Watts MJ, White M, Wintle BA, Yates C, Brook BW (2012) Plant extinction risk under climate change: are forecast range shifts alone a good indicator of species vulnerability to global warming? *Global Change Biology*, **18**, 1357–1371.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hastie T, Tibshirani R. (1987) Generalized additive models: Some applications. *Journal of the American Statistical Association*, **82**, 371–386.

He B, Lai T, Fan H, Wang W, Zheng H (2007) Comparison of flooding-tolerance in four mangrove species in a diurnal tidal zone in the Beibu gulf. *Estuarine, Coastal and Shelf Science*, **74**, 254-262.

Hongoh V, Berrang-Ford L, Scott ME, Lindsay LR (2012) Expanding geographical distribution of the mosquito, *Culex pipiens*, in Canada under climate change. *Applied Geography*, **33**, 53–62.

Jiménez-Valverde A, Lobo JM (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Ecologica*, **31**, 361–369.

Kearney MR, Wintle BA, Porter WP (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.

Keitt TH, Børnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography*, **25**, 616–625.

Kitaya Y, Jintana V, Piriyayotha S, Jaijing D, Yabuki K, Izutani S, Nishimiya A, Iwasaki M (2002) Early growth of seven mangrove species planted at different elevations in a Thai estuary. *Trees - Structure and Function*, **16**, 150-154.

Kremen C, Cameron A, Moilanen A (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320**, 222–226.

Leathwick J, Moilanen A, Francis M, Elith J, Taylor P, Julian K, Hastie T, Duffy C (2008) Novel methods for the design and evaluation of marine protected areas in offshore waters. *Conservation Letters*, **1**, 91–102.

Legendre P (1993) Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, **74**, 1659–1673.

Legendre P, Fortin MJ (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.

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

Lichstein JW, Simons TR, Shriner SA, Franzreb KE (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, **72**, 445–463.

Liu CR, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

McCullagh P, Nelder JA (1983) *Generalized Linear Models.* Chapman and Hall, London.

Meier ES, Lischke H, Schmatz DR, Zimmerman NE (2012) Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecology and Biogeography*, **21**, 164-178.

Meier ES, Lischke H, Schmatz DR, Zimmerman NE (2012) Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecology & Biogeography,* **21**, 164-178.

de Oliveira G, Araújo MB, Rangel TF, Alagador D, Diniz JAF (2012) Conserving the Brazilian semiarid (Caatinga) biome under climate change. *Biodiversity and Conservation*, **21**, 2913–2926.

Pagel J, Schurr FM (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.

Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Peterson AT, Ortega-Huerta MA, Bartley J, Sanchez-Cordero V, Soberon J, Buddemeier RH, Stockwell DRB (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature,* **416**, 626–629.

Phillips SJ, Dudik M, Schapire RE (2004*) A maximum entropy approach to species distribution modeling*. Proceedings of the 21st International Conference on Machine Learning, ACM Press, New York.

Rapacciuolo G, Roy DB, Gillings S, Fox R, Walker K, Purvis A (2012) Climatic associations of British species distributions show good transferability in time but low predictive accuracy for range change. *Plos One*, **7**, e40212.

Record S, Fitzpatrick MC, Finley AO, Veloz S, Ellison AM (2013) Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. *Global Ecology & Biogeography*, **22**, 760-771.

Revermann R, Schmid H, Zbinden N, Spaar R, Schroder B (2012) Habitat at the mountain tops: how long can Rock Ptarmigan (*Lagopus muta helvetica*) survive rapid climate change in the Swiss Alps? A multi-scale approach. *Journal of Ornithology*, **153**, 891–905.

R Core Development Team (2012) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Ridgeway G (2006) Generalized boosted regression models. *Documentation on the R package "gbm", version 1.5-7*, http://www.i-pensieri.com/gregr/gbm.shtml

Rojas-Soto OR, Sosa V, Ornelas JF (2012) Forecasting cloud forest in eastern and southern Mexico: conservation insights under future climate change scenarios. *Biodiversity and Conservation*, **21**, 2671–2690.

Santika, T. & Hutchinson, M.F. (2009) The effect of species response form on species distribution model prediction and inference. *Ecological Modelling*, **220**, 2365-2379.

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistics society, series B*, **36**, 111–147.

Swanson AK, Dobrowski S, Finley AO, Thorne JH, Schwartz MK (2013) Spatial regression methods capture prediction uncertainty in species distribution model projections through time. *Global Ecology and Biogeography*, **22**, 242-251.

Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Tarkhnishvili D, Gavashelishvili A, Mumladze L (2012) Palaeoclimatic models help to understand current distribution of Caucasian forest species. *Biological Journal of the Linnean Society*, **105**, 231–248.

Temby O (2009*) Project report for aerial LiDAR survey of Darwin and surrounds*. Fugro Spatial Solutions, Perth.

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, de Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL, Williams SE (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

Thuiller W, Albert C, Araújo MB, Berry PM, Cabeza M, Guisan A, Hickler T, Midgley GF, Paterson J, Schurr FM, Sykes MT, Zimmermann NE (2008) Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.

Williams D, Wolanski E, Spagnol S (2006) *Hydrodynamics of Darwin Harbour*, in The environment in Pacific Harbours (ed. by E. Wolanski). The Hague, The Netherlands.

Wintle BA, Bardos DC (2006) Modeling species-habitat relationships with spatially autocorrelated observation data. *Ecological Applications*, **16**, 1945–1958.

Wintle BA, Bekessy SA, Keith DA, van Wilgen BW, Cabeza M, Schroder B, Carvalho SB, Falcucci A, Maiorano L, Regan TJ, Rondinini C, Boitani L, Possingham HP (2011) Ecological-economic optimization of biodiversity conservation under climate change. *Nature Climate Change*, **1**, 355–359.

Zurell D, Grimm V, Rossmanith E, Zbinden N, Zimmermann NE, Schroder B (2012) Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps. *Ecography*, **35**, 590–603.

Table 1. Spatial autocorrelation indicated by Moran's Index values from model residuals from environment-only (Env-only), autologistic and RAC (residuals autocovariate) models, for the simulated organism (Snouter) and three mangrove communities (each dominated by one of *Ceriops tagal*, *Rhizophora stylosa* or *Sonneratia alba*), for a neighbourhood size of 3 x 3 cells.

	Env-only	Autologistic	RAC
Snouter (mean and s.e. from 10 simulations)	0.375 ± 0.06	-0.029 ± 0.01	-0.028 ±0.01
Ceriops tagal	0.322	-0.077	-0.082
Rhizophora stylosa	0.309	-0.087	-0.073
Sonneratia alba	0.356	-0.010	-0.036

Table 2. Model performance metrics shown for three modelling approaches: environment-only, autologisitic and RAC (residuals autocovariate) models for (a) Snouter, for 10 realizations of the simulated dataset; and (b) mangrove communities dominated by *Sonneratia alba*, *Rhizophora stylosa* or *Ceriops tagal*, for Wood's Inlet, northern Australia. AUC is area under curve. s.e. indicates standard error.

(a) Snouter

Model	Explanatory variables	AUC ± s.e.	% deviance explained ± s.e.
Environment	rain, djungle	0.68 ± 0.02	2.44 ± 0.36
Autologisitic	rain, djungle, autocov	0.95 ± 0.01	57.2 ± 2.8
RAC	rain, djungle, RAC	0.95 ± 0.01	56.9 ± 2.8

(b) Mangrove communities

Model	Explanatory variables	AUC ± s.e.	% deviance explained ± s.e.
Sonneratia alba			
Environment	hydroperiod, salinity, slope	0.99 ± 0.00	71.4 ± 0.4
Autologistic	hydroperiod, salinity, slope, autocov	0.99 ± 0.00	90.6 ± 0.3
RAC	hydroperiod, salinity, slope, RAC	0.99 ± 0.00	86.5 ± 0.3
Rhizophora stylosa			
Environment	hydroperiod, salinity, slope	0.92 ± 0.00	46.9 ± 0.9
Autologistic	hydroperiod, salinity, slope, autocov	0.98 ± 0.00	75.2 ± 0.6
RAC	hydroperiod, salinity, slope, RAC	0.98 ± 0.00	74.0 ± 0.7
Ceriops tagal			
Environment	hydroperiod, salinity, slope	0.90 ± 0.00	43.6 ± 0.06
Autologistic	hydroperiod, salinity, slope, autocov	0.98 ± 0.00	74.1 ± 0.7
RAC	hydroperiod, salinity, slope, RAC	0.98 ± 0.00	73.2 ± 0.8

Figure 1. Correlograms indicating spatial autocorrelation indicated by Moran's Index (y-axis) at different distances (the number of 25x25 m grid cells), for the raw observational data of mangrove communities dominated by *Sonneratia alba*, *Rhizophora stylosa* or *Ceriops tagal*, and the residuals from environment-only, autologistic and residuals autocovariate (RAC) models. Of the 17250 cells in the study area 1000 were sampled for these correlograms.

Figure 2. Partial plots showing the relationship between the probability of occurrence across a gradient in hydroperiod (duration of inundation as a percent of time), for three mangrove communities characterized by *Ceriops tagal*, *Rhizophora stylosa* or *Sonneratia alba*, for three modeling approaches: environment-only, RAC (residuals autocovariate) and autologistic. Black line indicates the mean, and grey lines the 95% confidence intervals based on bootstrapping the dataset with 17250 observations, 1000 times. Relative variable importance (R.I.) shown on top right corner of each panel indicates the proportion of times the variable hydroperiod was selected during the boosted regression tree (BRT) model fitting.

Figure 3. The model forecasts for the simulated organism, Snouter, for the initial level of rain and the six scenarios of increasing rain, provided by the environment-only, autologistic and residuals autocovariate (RAC) models. (a) the mean (± standard errors) of the expected number of cells occupied; and (b) the mean difference in the expected number of cells occupied.

Figure 4. Observed and modeled distribution of mangrove communities dominated by either *Ceriops tagal*, *Rhizophora stylosa* or *Sonneratia alba*, shown for a portion of Woods Inlet (2

by 2 km), Darwin Harbor, northern Australia. The observed distributions were derived from maps produced by Brocklehurst and Edmeades (1996) based on field surveys and aerial photography. The modeled distributions are shown for environment-only (env-only), residuals autocovariate (RAC) and autologistic models. Grey scale: black high probability, white low probability.

Figure 5. The mean expected number of cells (out of a total of 17250 in the study area) occupied by each of the three mangrove communities dominated by either (a) *Sonneratia alba*, (b) *Rhizophora stylosa* and (c) *Ceriops tagal*, for environment-only, autologistic and residuals autocovariate (RAC) models, and each sea level increment (0, 0.5 and 1 m), for Woods Inlet, Northern Territory, calculated from 1000 bootstrapped samples. Standard errors ranged from 0.28-2.47 (refer to Appendix S4). Values for *R. stylosa* and *C. tagal* are represented on the same scale (x-axis). Forecasts from the environment-only models and models accounting for spatial autocorrelation diverge with successive increments in sea level.