

TECHNOLOGICAL ADVANCES AT THE INTERFACE BETWEEN ECOLOGY AND STATISTICS

Multi-site generalised dissimilarity modelling: using zeta diversity to differentiate drivers of turnover in rare and widespread species

Guillaume Latombe^{*1} , Cang Hui^{2,3} and Melodie A. McGeoch¹

¹School of Biological Sciences, Monash University, Melbourne, Vic. 3800, Australia; ²Centre for Invasion Biology, Department of Mathematical Sciences, Stellenbosch University, Matieland 7602, South Africa; and ³African Institute for Mathematical Sciences, Cape Town 7945, South Africa

Summary

1. Generalised dissimilarity modelling (GDM) applies pairwise beta diversity as a measure of species turnover with the purpose of explaining changes in species composition under changing environments or along environmental gradients. Beta diversity only captures turnover across pairs of sites and, therefore, disproportionately represents turnover in rare species across communities. By contrast, zeta diversity, the average number of shared species across multiple sites, captures the full spectrum of rare, intermediate and widespread species as they contribute differently to compositional turnover.

2. We show how integrating zeta diversity into GDMs (which we term multi-site generalised dissimilarity modelling, MS-GDM), provides a more information rich approach to modelling how communities respond to environmental variation and change. We demonstrate the value of including zeta diversity in biodiversity assessment and modelling using BirdLife Australia Atlas data. Zeta diversity values for different numbers of sites (the order of zeta) are regressed against environmental differences and distance using two kinds of regressions: shape constrained additive models and a combination of I-splines and generalised linear models.

3. Applying MS-GDM to different orders of zeta revealed shifts in the importance of environmental variables in explaining species turnover, varying with the order of zeta and thus with the level of co-occurrence of the species and, by extension, their commonness and rarity. In particular, precipitation gradients emerged as drivers in the turnover of rare species, whereas temperature gradients were more important drivers of turnover in widespread species.

4. Appreciation of the factors that drive compositional turnover across multiple sites is necessary for accommodating the full spectrum of compositional turnover across rare to common species. This extends beyond understanding drivers for pairwise beta diversity only. MS-GDM provides a valuable addition to the toolkit of GDM, with further potential for survey gap analysis and prediction of species composition in unsampled sites.

Key-words: alpha diversity, beta diversity, compositional turnover, generalised additive models, generalised dissimilarity modelling, I-splines, richness, shape constrained additive models

Introduction

Understanding the drivers of patterns in species diversity is necessary to better protect regional diversity and ecosystem function, implement sound conservation management actions and prevent further species loss and assemblage homogenisation (Socolar *et al.* 2016). Modelling species richness from remotely mapped environmental variables is by far the most widely used approach to quantify the relationships between biodiversity and its environmental drivers (Ferrier 2002). However, local species richness, or alpha diversity, although comparatively easy to estimate, provides an incomplete picture of biodiversity and of species richness at regional scales, i.e.

gamma diversity, which is more difficult to assess (Socolar *et al.* 2016). To better quantify gamma diversity, it is therefore crucial to understand the drivers of species turnover, usually measured as turnover in species composition between pairs of sites, i.e. beta diversity (Jost 2007; Chao, Chiu & Hsieh 2012).

Generalised dissimilarity modelling (GDM) is an approach that enables the nonlinear modelling of pairwise beta diversity from differences in environmental conditions and distance between sites (Ferrier *et al.* 2007). In combination with the modelling of alpha diversity, GDM has been shown to be useful for modelling the species composition of large regions where exhaustive sampling is impossible, allowing for the interpolation of species assemblages across an entire region from limited sampling data, with potential applications for the management of such regions (Mokany *et al.* 2011). The approach

*Correspondence author. E-mail: guillaume.latombe@monash.edu

is also valuable for a wide range of other applications, such as survey gap analyses (Ferrier 2002; Funk, Richardson & Ferrier 2005; Ashcroft *et al.* 2010; Rose *et al.* 2016), climate-change impact assessment (Ferrier, Harwood & Williams 2010; Prober *et al.* 2012), and for visualising spatial patterns in community composition (Ferrier *et al.* 2004; Thomassen *et al.* 2010; Leitão *et al.* 2015). Rather than computing species turnover from separate predictions of the distribution of individual species (e.g. De'Ath 2012), approaches such as GDM are especially useful when extensive monitoring of all individual species has practical limitations, as for highly diverse biological groups with high turnover and encompassing multiple biotic interactions (Ferrier *et al.* 2007).

However, limiting the analysis of spatial compositional turnover in biodiversity to measures of alpha and pairwise beta diversity fails to comprehensively capture information on turnover in species assemblages (Hui & McGeoch 2014). When more than two sites or assemblages are involved in a study, the average of pairwise beta diversity metrics, such as Jaccard or Bray–Curtis indices, has been used (Jost, Chao & Chazdon 2011). For turnover across three sites, metrics based on pairwise beta diversity do not capture the number of species shared exclusively by pairs of sites, nor the number of species shared by all three sites. Multiple-site metrics have therefore been developed to quantify the heterogeneity in assemblage composition (Diserud & Ødegaard 2007; Baselga 2013; Ricotta & Pavoine 2015), but these measures have shortcomings, such as inference problems due to the averaging of non-independent pairwise values, and difficulty of interpretation for large number of assemblages (Hui & McGeoch 2014). Zeta (ζ) diversity, which measures the number of species shared by any number of assemblages or sites, has been proposed *inter alia* as a means to overcome these limitations (Hui & McGeoch 2014). As the number of sites increases (the order of zeta, noted as ζ_i for i different sites in the following), the number of species that they share necessarily declines. Rare species, shared by a small number of sites, therefore disproportionately contribute to low orders of zeta diversity (and to beta diversity), whereas widespread species, shared by a large number of sites, contribute to higher orders of zeta diversity (note that the characteristic of a species to be rare or widespread depends on the grain and extent considered). Zeta diversity is thus a metric that captures all incidence-based diversity components produced by assemblage partitioning, and can be used to generate formulae for a range of community patterns.

When applied to pairwise beta diversity-based measures of compositional turnover, GDM predominantly explains the relationship between environmental variation and turnover in the rare species composition of a community. To better understand variation in compositional similarity, understanding the drivers of turnover in rare, intermediate and widespread components of the community is required. Given the recent acknowledgement that the value of common species for ecosystem function may have been underestimated (Gaston 2010) and that being common is not a straightforward quality but can be characterised in a multidimensional fashion (McGeoch & Latombe 2016), disentangling the

turnover of species with different levels of occurrence and co-occurrence is essential for conservation and for predicting the impact of land use and climate change on biodiversity. Here, we extend the concept of GDM to incorporate zeta diversity, and therefore to understand the importance of environmental gradients and spatial distance in explaining the compositional turnover of the whole spectrum of species, from rare (low occurrence) to widespread (high occurrence). We therefore expect the relative contribution of each variable to the explained variation in species turnover to vary with the increase in the order of zeta diversity. This is because, as the order of zeta increases, we gradually remove the contribution of rare species to the variation of observed turnover, and the contribution of different niche characteristics are likely to emerge for rare vs. widespread species. However, since the variability in species composition over different sites increases as we consider more sites, we also expect a decline in the predictive power of multi-site generalised dissimilarity modelling (MS-GDM) for higher orders of zeta diversity.

Zeta diversity is the number of species in common among sites, and is a measure of similarity, rather than dissimilarity. However, its application to GDM is essentially an extension of GDM's scope to encompass multi-site measures of turnover. We therefore call this MS-GDM. We apply MS-GDM to the distribution of bird species occurrences across Australia. To explore the importance of different environmental gradients for turnover, we apply and compare two different techniques to perform MS-GDM: (i) nonlinear decreasing regression of environmental differences and distance between sites, introducing the use of shape constrained additive models (SCAM; Pya & Wood 2015) for this purpose; and (ii) decreasing generalised linear modelling of the difference between environmental variables transformed with I-splines, a class of monotonic spline functions classically used in GDM (Ferrier *et al.* 2007). For a study area as large as Australia, the I-spline-based MS-GDM is likely to perform better than the SCAM-based MS-GDM. This is because the rate of change in species composition is likely to vary along wide environmental gradients. Although the bird occurrence dataset is used primarily for illustrating the methods rather than fully understanding turnover in bird species composition in Australia, MS-GDM highlights differences between different levels of co-occurrence and therefore of rarity and commonness, a distinction that has implications for managing biodiversity.

Materials and methods

BIRD OCCURRENCE DATA

We used bird survey data (presence-only) from the BirdLife Australia Atlas of Australian Birds, from 1998 to 2013 covering all states and territories of Australia (Barrett *et al.* 2003). The dataset comprises 761 species, among which we selected the component consisting of 258 terrestrial, resident species (see Appendix S1, Supporting Information for details). The Australian continent and Tasmania were divided into 25×25 km and 100×100 km cells to assess the impact of grain on

the analyses. Each cell was considered as a site in the analyses of zeta diversity, and was characterised by its xy coordinates and its species composition. Cells whose observed richness was likely to underestimate real richness were removed to limit the impact of false absences on the analyses (Fig. 1; Appendix S1).

ENVIRONMENTAL VARIABLES

Fourteen environmental variables that have previously been shown to explain bird community structure were considered, including the percentage of land use types per unit area and climate-related variables (Jetz, Wilcove & Dobson 2007), human density (Chown *et al.* 2003; Gaston & Evans 2004), and elevation (Rahbek 1997). Although elevation is not a proximal predictor of species distribution, it is commonly used as an effective surrogate for other climate variables (Elith & Leathwick 2009) and more complex variables, such as food availability (Remonti, Balestrieri & Prigioni 2009) or vegetation distribution (Dirnböck *et al.* 2002), that are more difficult to measure directly. Variables in addition to the ones selected here may explain bird diversity, such as vegetation structure and food resources (Cody 1981; Ferger *et al.* 2014), but we limited this assessment to this subset of variables that are adequate for the purpose of demonstrating MS-GDM, its application, and the insights it provides.

Maps with a 25×25 km and a 100×100 km cell resolution were computed for all variables (Figs B1–B3, Appendix S2). To account for collinearity between variables and non-independence between the land use variables, we computed the correlation matrix between all 14 variables at the 25 km grain, and removed variables so that Pearson's correlation values r between the remaining ones were $-0.5 < r < 0.5$. As a result, only seven variables were retained: the area of conservation and natural habitats (Natural) per map cell (expressed as a percentage over the whole area of the cell), the area of production from irrigated agriculture and plantations (Irrigated) per cell, the area of water bodies (Water; including estuary/coastal water, marsh/wetlands, rivers, lakes and reservoirs/dams) per cell, as well as human density, mean temperature, average monthly precipitation and elevation (Table B1, Appendix S2). The human density was then transformed to compute its reciprocal, the area-per-person (bounded between 0 and 1), to reduce the skewness of the distribution (Fig. B2, Appendix S2).

MULTI-SITE GENERALISED DISSIMILARITY MODELLING

Generalised dissimilarity modelling is typically used to explain the value of beta diversity for pairs of sites, measured as any diversity index bounded by 0 and 1 (such as Bray–Curtis or Jaccard dissimilarity indices), based on the difference in environmental variables and spatial

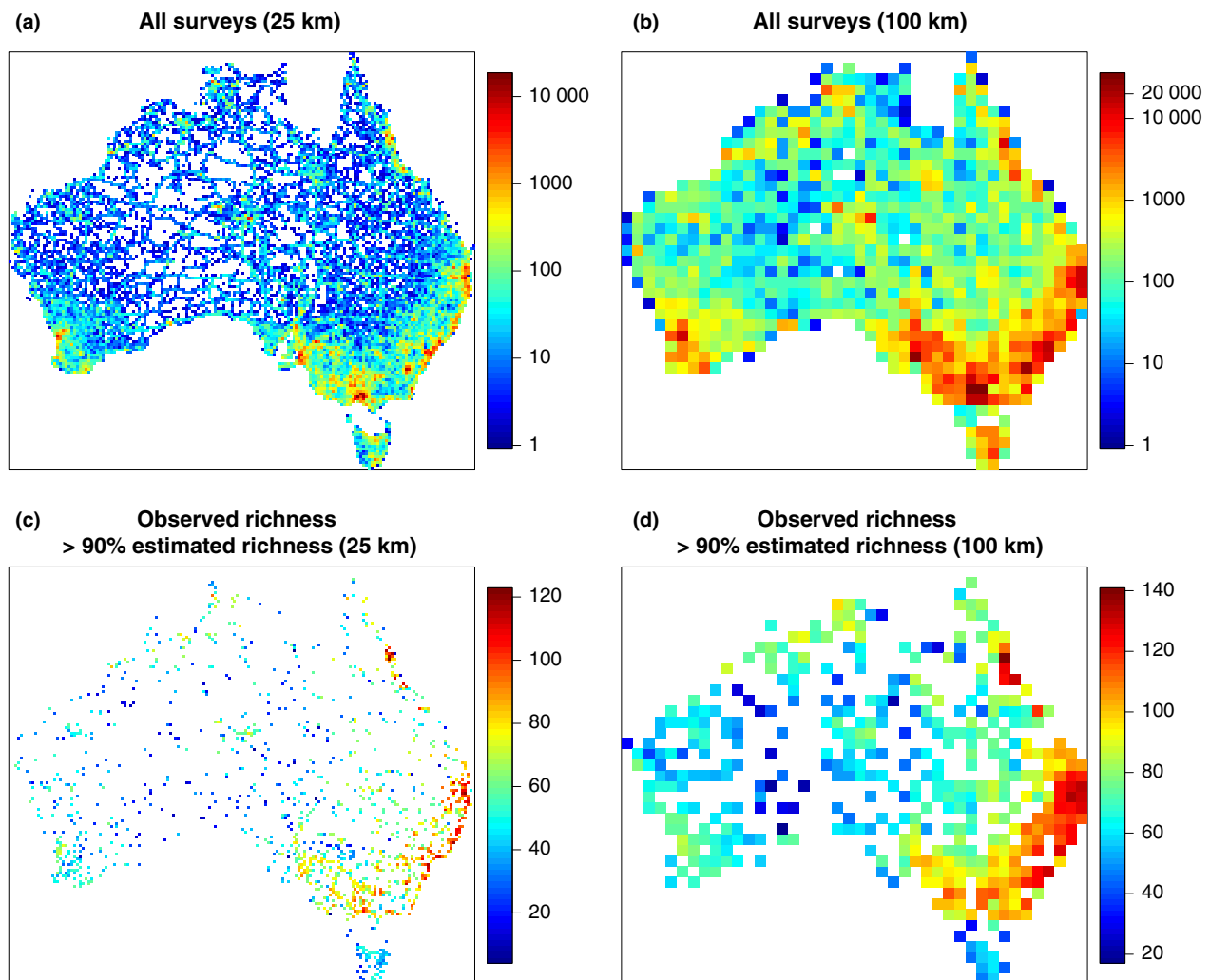


Fig. 1. Maps of the number of surveys (log-scaled) at (a) the 25 km grain and (b) the 100 km grain, and maps of the cells with at least five surveys and within 90% estimated Chao 2 richness at (c) the 25 km grain and (d) the 100 km grain (see Appendix S1 for details).

distance between sites, using regression models. In the same way, MS-GDM extends GDM to explain the value of zeta diversity for n -number of sites (or zeta order) based on environmental difference and spatial distance between them using regression models. As we show in the following, since ζ_2 is a measure of beta diversity, MS-GDM applied to ζ_2 is equivalent to GDM, providing some transformation to keep the zeta value between 0 and 1. However, applying MS-GDM to higher orders of zeta provides additional information on the drivers of species turnover as we show below.

SCAM-based MS-GDM

Like GDM, MS-GDM must be able to accommodate two constraints with respect to the relationship between zeta diversity and environmental gradients or distance. First, MS-GDM is based on the assumption that, if species composition is driven by environmental and spatial variables, the number of species in common between a given number of sites necessarily decreases monotonically with the environmental gradient and spatial distance (Ferrier *et al.* 2007). In practice, however, it is possible to observe local increases in zeta diversity with environmental gradients or spatial distance. For example, a strong, spatially periodic environmental gradient can lead to observed spatial periodicity in species turnover (Nekola & White 1999). Increases in zeta diversity with increasing environmental gradients or spatial distance are therefore likely to represent unexplained relationships between the environment and species composition, and cannot be used to predict species turnover. The regressions between zeta diversity and environmental gradients or spatial distance must therefore be forced to have negative coefficients.

Second, similar to measures of beta diversity, such as Jaccard or Bray–Curtis indices, that are bound by 0 and 1, the number of species in common between sites (zeta) for a limited spatial extent is necessarily constrained between zero and the total number of species (Faith, Minchin & Belbin 1987; Ferrier *et al.* 2007). Building on the first constraint of monotonic decrease in zeta diversity with environmental gradients and spatial distance (zeta decay *sensu* Hui & McGeoch 2014), zeta diversity will decline in a curvilinear fashion to asymptote at a lower limit of zero (or in some cases a value higher than zero due to a few widespread species) as environmental gradients and spatial distance increase.

To accommodate these two constraints, we demonstrate how MS-GDM can be performed using the recently introduced semi-parametric SCAM (Pya & Wood 2015), using the package SCAM (Pya 2014) in R (R Core Team 2013). SCAMs are similar to generalised additive models (GAM; Hastie & Tibshirani 1990), but use mildly nonlinear extensions of P-splines to incorporate constraints in the shape of the GAM. In particular, the SCAM package enables the imposition of a monotonic decrease on the P-splines, an essential feature for satisfying the two constraints described above. Since SCAMs, like GAMs, are semi-parametric, they potentially offer more flexibility and accuracy than generalised linear models (GLMs), and it is worth testing their application for MS-GDM, although their outputs can be less straightforward to interpret and may be more prone to overfitting problems than GLMs (Hastie & Tibshirani 1990).

A SCAM-based MS-GDM for ζ_2 can be expressed as follows:

$$\zeta_2 = \sum_p f_p(|x_{p1} - x_{p2}|) \quad \text{eqn 1}$$

where p is an environmental variable or the optional spatial location, x_{p1} is the value of variable x_p at location 1, and f_p is a SCAM. For ζ_n , $n > 2$, eqn (1) becomes:

$$\zeta_n = \sum_p f_p(g(\{|x_{pi} - x_{pj}|, [i, j] \in N^2, i \neq j\})) \quad \text{eqn 2}$$

where g is a function to combine environmental differences and distance, N is a subset of n sites from all the sites and $g(\{|x_{pi} - x_{pj}|, [i, j] \in N^2, i \neq j\})$ is therefore a measure of dispersion of x_p . In theory, any measure of dispersion can be used. The most intuitive measures are the mean and the maximum difference or distance. As the order of zeta increases, the measure of dispersion for any set $\{|x_{pi} - x_{pj}|, [i, j] \in N^2, i \neq j\}$ will converge to the value for the overall distribution of $\{|x_{pi} - x_{pj}|, [i, j] \in N_{\text{tot}}^2, i \neq j\}$, where N_{tot} is the set of all sites. As a result, all measures of dispersion will converge to a single value, which will make the regression of eqn (2) impossible. The convergence rate will nonetheless differ between the different measures of dispersion. We tested the impact of using the mean and the maximum functions on the output of MS-GDM. Since the environmental variables and the locations have a limited range, the maximum converged faster than the mean due to the chance of selecting locations on the edges of the study area increasing with the order of zeta (see Figs B3 and B4, Appendix S2, for distances at the 100 km grain). This produces singularity problems during the regression of eqn (2) for high orders of zeta at the 100 km grain. Since results were similar for both methods for low orders of zeta, only the results for the mean are presented here.

The number of species in common between any number of sites is necessarily bounded by the overall number of species among all sites S , and ζ_n can be divided by S_n in eqn (2), which results in a proportion. We therefore used a binomial variance distribution function with a logit link for the computation of the SCAM in eqn (2) (Ferrier *et al.* 2007). Note that different transformations of zeta diversity can be used (such as dividing by the average richness or the local richness), which assume different mean–variance relationships in the data, a common problem in distance-based analyses (Warton, Wright & Wang 2012). Although beyond the scope of this paper, understanding the effects of different transformations on the analyses will be important for future developments of MS-GDM (e.g., Heino *et al.* 2013).

I-spline-based MS-GDM

Shape constrained additive models-based MS-GDM assumes that a specific environmental gradient implies the same compositional dissimilarity, irrespective of the value of the environmental variables. From this perspective, SCAM is therefore similar to other methods based on linear models such as variation partitioning with function *varpart* from the VEGAN R package (Oksanen *et al.* 2016) or the DistLM function in PRIMER/PERMANOVA+ (Clarke & Gorley 2015), but with much greater flexibility. However, as pointed out by Ferrier *et al.* (2007), the rate of compositional turnover may potentially change along an environmental gradient. For example, small increases in temperature may be more detrimental to species near their upper thermal tolerances at a range edge compared with larger temperature changes towards the centre of their thermal tolerance performance curves (Calosi, Bilton & Spicer 2008). Simply applying a GLM or a SCAM, as described above, to environmental difference cannot accommodate such gradients when they occur.

Following Ferrier *et al.* (2007), to allow for this nonlinearity to structure the model, the nonlinear functions must apply to the environmental variables, not to the environmental difference between sites, and we transformed the environmental data using a linear combination of I-splines (Ramsay 1988):

$$I_p(x_p) = \sum_{k=1}^{s_I} a_{pk} I_{pk}(x_p) \quad \text{eqn 3}$$

where x_p is a value of the environmental variable p , I_{pk} is the k th spline of the I-spline family composed of s_I splines for the variable p , and a_{pk} is a non-negative coefficient, ensuring the monotonic increase in I_p . s_I corresponds to the sum of the order of the spline, and the number of interior knots, both determined by the user. Here we used the same default parameters as in GDM, i.e. an order of 2 and 1 knot located at the median, which generates 3 I-splines per variable p . Allowing more than one I-spline per environmental variable enables a monotonic yet potentially flexible fitted function.

Still following Ferrier *et al.* (2007), the a_{pk} were determined by solving the following equation in a GLM with a log link function and a binomial variance distribution function. Note that the log link function is equivalent for similarity measures to the link function recommended by Ferrier *et al.* (2007) for dissimilarity measures.

$$\log(\zeta_n/S) = - \sum_p \sum_{k=1}^{m_{S_I}} a_{pk} g(\{|I_{pk}(x_{pi}) - I_{pk}(x_{pj})|, [i,j] \in N^2, i \neq j\}) \quad \text{eqn 4}$$

where S is the total richness over the study and is used to rescale the values between $[0,1]$, as required by the binomial variance distribution function (but, as for SCAM-based MS-GDM, different transformations are possible). Forcing negative coefficients was achieved using a non-positive least-square regression from the `nnls` package (Mullen & van Stokkum 2012) within the iterative re-weighted least-square algorithm implemented in the `GLM2` package (Marschner 2014) in `R` (R Core Team 2013).

When sites are located along a transect, the spatial location along the transect can be treated as an environmental variable, and the I-splines directly applied to the location using eqn (4). This is useful if the species composition is expected to change in a nonlinear fashion along the transect. However, for two-dimensional coordinates, it is not possible to directly apply the I-splines to the site locations. One possibility is to apply the I-spline on the X and Y coordinates separately and therefore uses two measures of distance in eqn (4). However, the X and Y coordinates may not be appropriate, as the species composition may change along many local axes in different directions over the study area (i.e. being anisotropic). Tensor product splines (Wood 2006) may accommodate this constraint, but their combination with I-splines is not straightforward. Following Ferrier *et al.* (2007), when spatial distance is included in the MS-GDM, the I-splines were therefore simply applied directly to the distance D_{ij} , and eqn (4) becomes:

$$\log(\zeta_n/S) = - \left(\sum_p \sum_{k=1}^{m_{S_I}} a_{pk} g(\{|I_{pk}(x_{pi}) - I_{pk}(x_{pj})|, [i,j] \in N^2, i \neq j\}) \right) - \left(\sum_{k=1}^{m_{S_I}} a_k g(\{|I_{Dk}(D_{ij}), [i,j] \in N^2, i \neq j\}) \right) \quad \text{eqn 5}$$

Note that the difference of the I-splines was used in a GLM rather than a SCAM, because it would not have been possible to retrieve eqn (3) otherwise, as SCAMs are not distributive ($f(x - y) \neq f(x) - f(y)$), and the output would not be interpretable.

REGRESSION, SIGNIFICANCE TESTING AND EVALUATION

Multi-site generalised dissimilarity modelling was performed for the two grains (25×25 km and 100×100 km grains), using the SCAM-

based and the I-spline-based versions for orders of zeta diversity 2, 3, 4, 5, 10, 25, 50 and 100. To assess the impact of site selection on the results, we also performed the analyses for a smaller extent (see Appendix S4 for details). Changing the grain and extent only had limited impact on the results, and only results for the 100 km grain over Australia are reported here, whereas the outputs of the additional analyses are reported in Appendix S4.

Eight predictors, i.e. the seven environmental variables described previously as well as distance, were used in the analyses. For the SCAM-based MS-GDM, the environmental differences and distance were divided by their maximum value prior to fitting the models so that the coefficients a_{pk} could be directly compared. The original values were divided by their maximum rather than rescaled between 0 and 1 because, assuming the data covers the whole extent of the study area and the whole range of environmental values, additional sites may lead to smaller, but not larger, environmental gradients or spatial distances between sites. For the I-spline-based MS-GDM, the I-splines already rescale the environmental values between 0 and 1, so no additional transformation was performed. The performances of the two techniques for the different datasets were assessed by computing Pearson's r^2 between the observed and predicted zeta values.

To confirm our analyses, we compared MS-GDM for pairs of sites with a GDM computed with the `GDM` package (Manion *et al.* 2016) in `R` (R Core Team 2013). For the results generated by the two methods to be comparable, we used the 'normalised' ζ_2 , i.e. dividing the number of species in common between two sites by the overall number of species over the same two sites, which is equal to the Jaccard similarity index (note that, by extension, MS-GDM can be applied to any multi-site measure of similarity, or even dissimilarity by changing the signs of the coefficients in the different equations). Accordingly, the GDM was computed for the Jaccard dissimilarity index, and we confirmed that the two analyses yielded the same coefficients with opposite signs for each variable (Appendix S3).

The significance of the different splines (for the SCAM- or the I-spline-based MS-GDM) can be assessed using permutation matrices, following the method described by Ferrier *et al.* (2007). However, in practice, the current implementation of MS-GDM is too slow to compute the significance of the variables for all orders of zeta. We therefore used the implementation from the `GDM` `R` package (Manion *et al.* 2016) to assess the significance of the variables for ζ_2 (Appendix S3). Incorporation of Bayesian bootstrap into GDM has also recently been proposed for assessing the uncertainty in GDM parameters and predictions (Woolley *et al.* 2016), and is of interest for future developments of MS-GDM.

For all MS-GDM analyses except the comparison with the GDM, 50 000 combinations of sites were randomly selected using Monte Carlo sampling. For the comparison with the GDM, we used all possible pairs of sites for each of the two grains and extents, to make sure that the results were comparable between the two analyses. All functions for performing the analyses are included in the `R` package `ZETADIV` (Latombe *et al.* 2016).

To make sure the analyses captured meaningful relationships between the environmental variables and zeta diversity, and to mitigate the inflated Pearson's r^2 values due to non-independence between site combinations, we also performed the analyses for 70% of the sites, and evaluated the predictions of zeta values for the different orders with the observed zeta values over the remaining 30% sites by computing Pearson's r^2 between the observed and predicted zeta values. Moreover, to completely remove the possibility that a site is included in multiple compositions, we further computed Pearson's r^2 between the observed

and predicted zeta values with each site belonging only to a single combination (i.e. sampling the sites without replacement), re-arranging the combinations 300 times, and reported the distribution of the r^2 values. Since the number of such independent combinations decreased as the order of zeta increases, the computation of Pearson's r^2 between the observed and predicted zeta values is unreliable for very high orders of zeta, and we only performed these analyses for orders 2–10.

Because simple patterns in local richness (alpha diversity) are a useful basis for interpreting patterns in turnover, we fitted a GAM to ζ_1 (alpha) and all environmental variables and spatial coordinates, to obtain a more complete picture of terrestrial bird species diversity over Australia and its relationship with the environment. The function *gam* in package *mgcv* (Wood 2011) in R (R Core Team 2013) was used with default parameters. The flexibility of GAMs can accommodate nonlinear relationships between richness and environmental variables (Hortal *et al.* 2013). We used a GAM instead of a SCAM for ζ_1 because species richness is expected to be positively correlated with variables such as temperature (Hawkins *et al.* 2003), but negatively with the area-per-person (Chown *et al.* 2003).

Results

RICHNESS (ζ_1): GAM

Pearson's r^2 between the observed and predicted species richness values was 79.24% at the 100 km grain. The spatial coordinates, area of natural habitat per cell, elevation, human density, mean temperature and precipitation were significant ($P < 0.05$), whereas the area of production from irrigated agriculture and plantations and the area of water bodies were not ($P > 0.05$). Low P -values should nonetheless be interpreted with caution for GAMs, because smoothing parameter uncertainty is neglected in their computation, and therefore underestimates them (Wood 2006). For example, the spline of the area of natural habitat is relatively flat, meaning that it has little effect on species richness. By contrast, as expected, the richness increases with the human density (i.e. decreases with the area-per-person), temperature and precipitation (Fig. 2).

EXPLAINING TURNOVER FROM ENVIRONMENTAL GRADIENTS: SCAM-BASED MS-GDM

Performance

The maximum Pearson's r^2 between the observed and predicted zeta values across all orders of zeta, all datasets, and all the locations was 41%, and corresponded to order 4 at the 100 km grain (Fig. 3a). The order of magnitude of the variance explained was similar for the evaluation data and the full datasets except for the 100 km grain over SE Australia. Compared to the variance computed over the evaluation data, the variance explained for the full dataset was slightly overestimated for orders >2 (Fig. D4, Appendix S4), which may explain why the variance explained is higher for orders 3 and 4 than for order 2 (Fig. 3a). The coefficients of the linear models forced through an intercept of zero between observed (predictor) and estimated (response) zeta values were all below 1 and higher than 0.8 for orders ≤ 10 (Fig. 3c). The SCAM-based

MS-GDMs therefore systematically and slightly underestimated species turnover for low orders of zeta. This consistent bias across zeta orders suggests that one or more additional environmental variables are important for driving species turnover. Decreasing the grain resulted in a loss of explanatory power (Fig. 3a) and in accuracy (Fig. 3c) of the SCAM-based MS-GDM. The variance explained and the coefficients were highest for the 100 km grain over SE Australia. Since this is the dataset with fewest sites and the distributions of the variance of the distributions of Pearson's r^2 between the observed and predicted zeta values for the evaluation dataset was high, these values suggest overfitting for this dataset, and the splines generated by the model should be interpreted with caution.

Relative importance of environmental variables

The splines generated by SCAM represent the relationship between environmental difference (independent of the raw values of the environmental variables) or spatial distance, and the number of species shared for that particular order of zeta (Fig. 4). The slopes of the splines therefore indicate the importance of each variable for turnover, whereas a flat spline (slope = 0) means that the variable does not explain turnover.

Overall, SCAM-based MS-GDM suggests that temperature, precipitation, the natural habitat area, the area of water bodies and distance were major drivers of species turnover for most orders of zeta relative to the other variables (Fig. 4). The SCAM splines were mostly linear, indicating that turnover in species composition tended to increase steadily as environmental conditions changed and with distance between sites.

As the order of zeta increases, natural habitat area explained more of the turnover compared to the other variables (Fig. 4). By contrast, the importance of distance decreased, whereas the importance of temperature and precipitation was constant.

ACCOMMODATING ENVIRONMENTAL GRADIENTS: I-SPLINE-BASED MS-GDM

Performance

Pearson's r^2 between the observed and predicted zeta values was higher for I-spline-based than for SCAM-based MS-GDM for all orders of zeta and all datasets. The maximum Pearson's r^2 between the observed and predicted zeta values across all orders of zeta, all datasets, and all the locations was 71%, and corresponded to order 4 at the 100 km grain (Fig. 3b). As in the SCAM-based MS-GDM, Pearson's r^2 between the observed and predicted zeta values was slightly overestimated for orders >2 with respect to the variance computed over the evaluation data (using 30% of the data), but the orders of magnitude were similar for the evaluation and the full datasets except for the 100 km grain over SE Australia (Fig. D8, Appendix S4). Like with the SCAM-based MS-GDM, the coefficients of the linear models forced through an intercept of zero between observed (predictor) and estimated

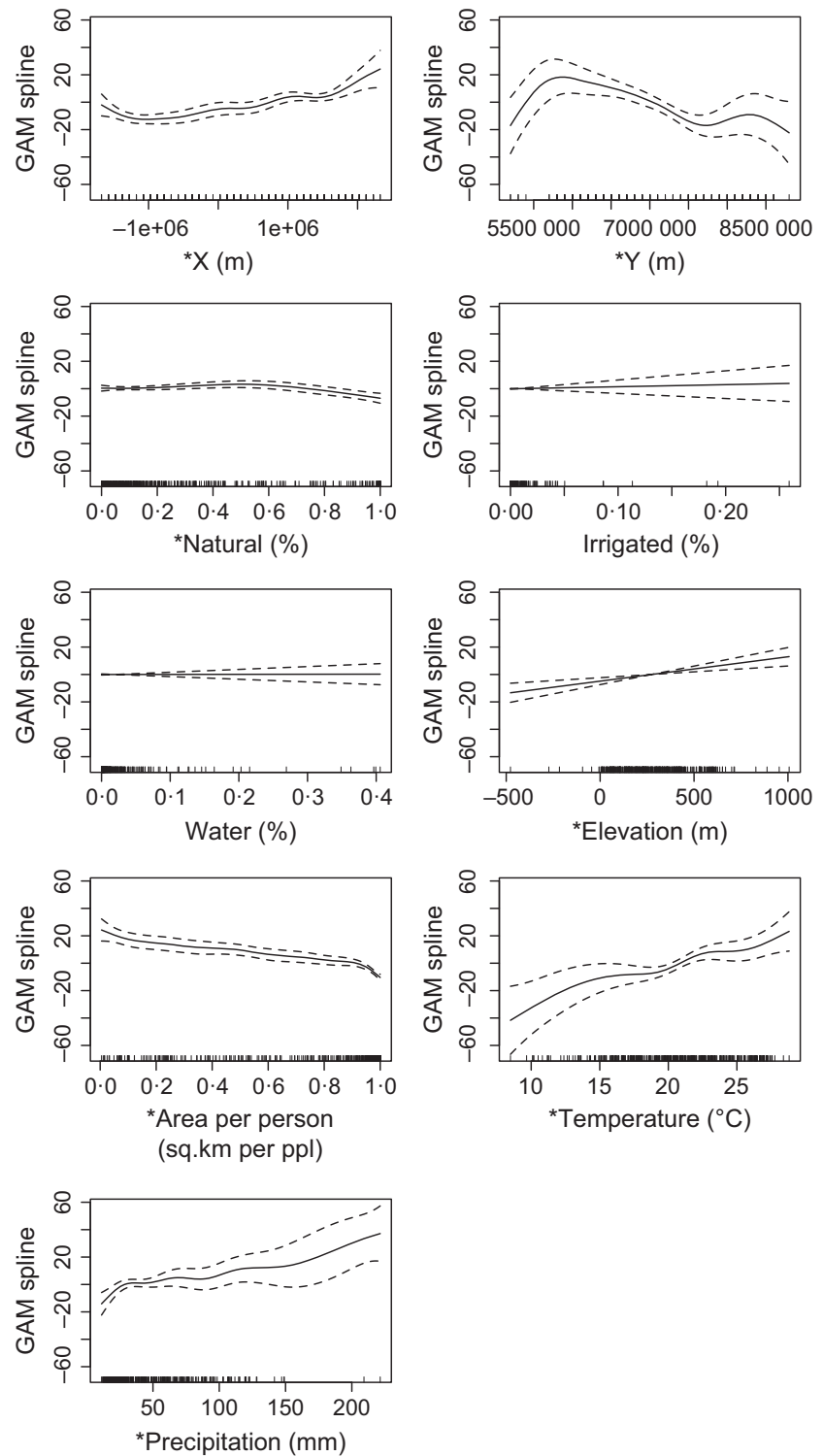


Fig. 2. Generalised additive model (GAM) splines describing the contribution of environmental variables to describing species richness (S , i.e. ζ_1) over the environmental range at the 100 km grain over Australia (see Figs D1–D3 in Appendix S4 for other grains and extents). The rug on the x -axis shows the distribution of the data. Significance is indicated by an * ($P < 0.05$). The model explained 79.24% of the variation in species richness.

(response) zeta values were lower than 1 and >0.8 , meaning that I-spline-based MS-GDMs also systematically but only slightly underestimated species turnover for low orders of zeta (Fig. 3d). This suggests the possible contribution of other variables in addition to those used here for predicting species turnover. The coefficients were nonetheless always closer to 1 than for the SCAM-based MS-GDM, meaning that I-spline-based MS-GDM was more accurate. Results also suggest overfitting for the 100 km grain over SE Australia.

Overall relative importance of environmental variables

The I-splines of each variable computed by eqn (3) with the coefficients obtained after fitting eqn (5) provide two useful pieces of information (Ferrier *et al.* 2007): (i) the maximum value of the spline indicates the importance of the variable relative to others for the same order in explaining zeta diversity, whereas (ii) the variation in slope indicates how the rate of compositional turnover changes along an environmental

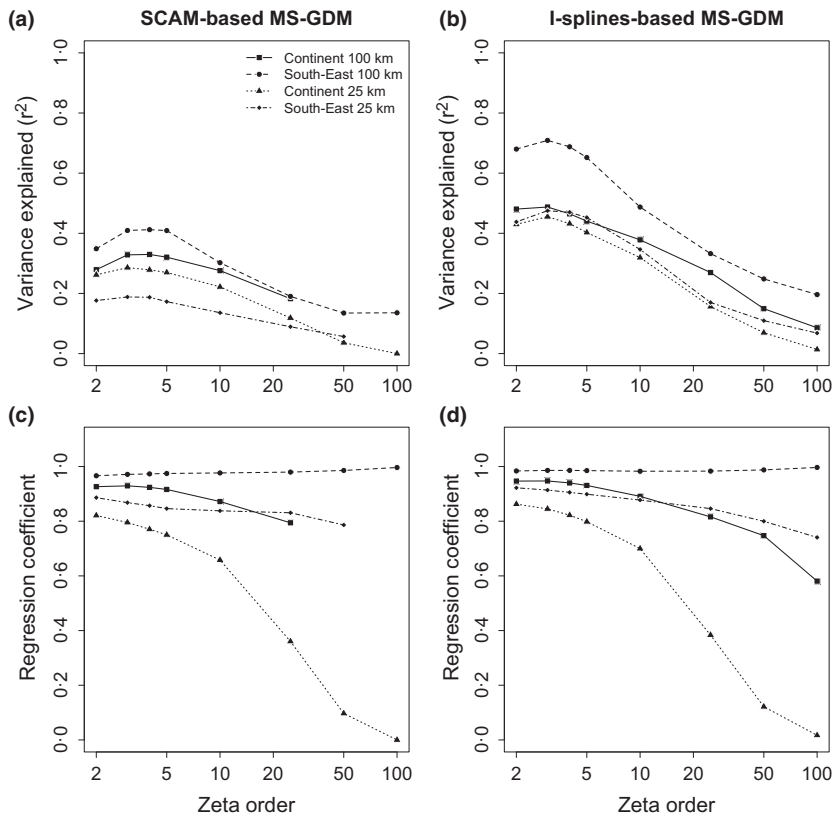


Fig. 3. Model performance (Pearson's r^2 between the observed and predicted zeta values) for the relationship between observed (predictor) and estimated (response) zeta values with an intercept of zero, computed over all four datasets for the different orders of zeta for (a) shape constrained additive models (SCAM)-based multi-site generalised dissimilarity modelling (MS-GDM) and (b) I-splines-based MS-GDM. Coefficients of a linear model with an intercept of zero between observed and predicted zeta values for (c) SCAM-based MS-GDM and (d) I-splines-based MS-GDM. Coefficient values lower than 1 indicate underestimation of the zeta values by the model. Convergence problems sometimes prevented computation of the SCAM for high orders in the SCAM-based MS-GDM.

gradient. For example, a sharper increase at low than at high precipitation indicates that changes in precipitation are most important for zeta diversity in dry areas (Fig. 5).

Overall, I-spline-based MS-GDM suggests that precipitation, temperature, natural habitat area, human density, elevation and distance are major drivers of species turnover for most orders of zeta (Fig. 5). As expected, the I-splines displayed important nonlinearities, indicating that the rate of compositional turnover changed along environmental gradients. The I-splines for precipitation and water bodies indicate that the rate of compositional turnover was high in dry areas. The rate of compositional turnover was also high near the extremes of the temperature range in Australia, but low in the middle of the range (Fig. 5). Changes in species composition occurred at low rather than high elevation, low human density (i.e. high area-per-person), large areas of natural habitat and either low or high distances.

Changes in the importance of variables with the order of zeta

Precipitation was the main driver of species turnover at low orders of zeta (Fig. 5). As the order increased, the importance of temperature increased, and temperature became the main driver of turnover for orders higher than 5 (i.e. the I-splines had the highest values). The importance of natural habitat area also increased with the order of zeta, although in a slower fashion, to be as important as precipitation for orders higher than 25. The importance of elevation increased for intermediate orders of zeta, but decreased for high orders. The presence of

water bodies was unimportant at low orders of zeta, but moderately important for higher orders, especially at order 25.

Discussion

By adapting the concept of GDM to zeta diversity, we show that the relative importance of the different environmental variables in explaining bird species turnover across Australia varies across the orders of zeta. Although MS-GDM does not solve all the shortcomings associated with the use of beta diversity, such as the averaging of non-independent pairwise values when characterising environmental differences between sites (eqns 2 and 4), it demonstrates the importance of using a more complete representation of compositional turnover than only pairwise dissimilarity to understand what is driving different components of biodiversity. Applying MS-GDM to ζ_2 , like GDM, aims to explain variation in species turnover for the whole community, as species shared by any pair of sites can be either common (more probable) or rare (less probable). As the order of zeta increases, MS-GDM explains variation in species turnover for more widespread species, since, by definition, rare species are not shared by many sites. Comparing the outputs of MS-GDM for a range of orders of zeta therefore provides a more comprehensive picture of species turnover. Comparing turnover in widespread species (high zeta orders) with turnover in the whole community has the potential to improve applications of the traditional GDM, such as the modelling of species composition. Although, as expected, the performance of MS-GDM decreased with the order of zeta, the performance

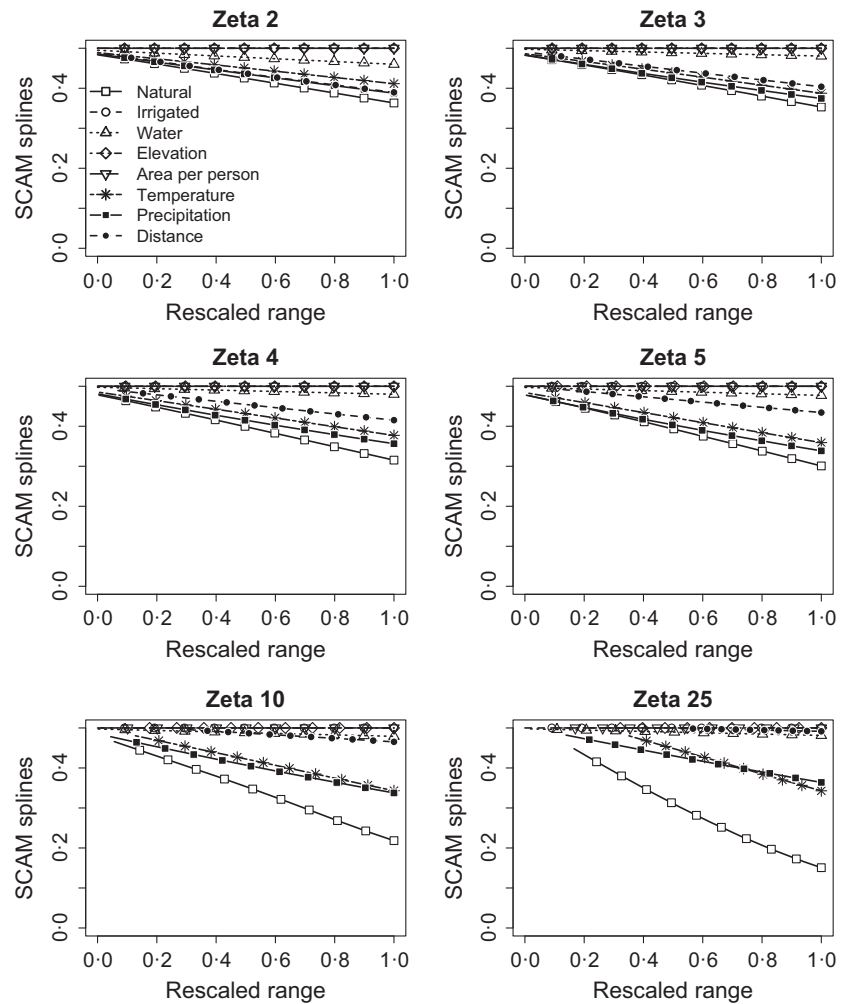


Fig. 4. Splines of the shape constrained additive models (SCAM) showing the contribution of the seven environmental variables and distance to explaining zeta diversity, for ζ_2 – ζ_{25} , at the 100 km grain over Australia. The splines represent changes in compositional turnover with environmental gradient and spatial distance. The steeper the slope of decline, the more species composition changes with the environment. All predictors are scaled between 0 and 1 for visualisation (rescaled range). See Table D4 in Appendix S4 for the scaling factors of all orders of zeta. The exact values of the scaling factors depend on the specific Monte-Carlo sample. Convergence problems prevented computation of the SCAM for orders 50 and 100.

remained satisfactory for a range of orders (Pearson's r^2 between the observed and predicted zeta values $>30\%$ and regression coefficient >0.8 for orders ≤ 10).

We used two statistical techniques to apply MS-GDM. The first makes use of the recent GAM under shape constraint, or SCAM (Pya & Wood 2015). Like GAM, SCAM is a semi-parametric technique and does not require a specific response curve to transform the predictor variables (Hastie & Tibshirani 1990). SCAM therefore provides a high degree of flexibility for inferring the relationship between zeta diversity and environmental gradients and spatial distance. However, SCAM does not accommodate changes in the rate of compositional turnover along an environmental gradient. To do so, following Ferrier *et al.* (2007), the environmental variables were transformed using I-splines before applying a GLM. Changes in the rate of compositional turnover along environmental gradients are characterised by nonlinear I-splines, and such nonlinearities are more likely to appear for large areas with wide ranges of environmental values. The areas covered in our analyses are large for all datasets and the I-splines displayed nonlinearities. As a result, Pearson's r^2 between the observed and predicted zeta values was always higher for the I-spline-based MS-GDM than for the SCAM-based MS-GDM, for both the full dataset

and the evaluation data. SCAM-based MS-GDM may provide better results for smaller study areas, as developed below.

Moreover, since for low orders of zeta, changes in precipitation are especially important for dry areas, i.e. a small range of values over the whole range of precipitation over Australia, SCAM-based MS-GDM underestimated the importance of precipitation for species turnover due to its inability to accommodate such nonlinearities (Figs 4 and 5). Using I-splines, nonetheless increases the number of parameters, and one should control for the number of knots and the order of the splines to avoid overfitting (Elith *et al.* 2006). Here, risks of overfitting were limited by the fact that we used only three splines. Low chance of overfitting was supported by the predictions on the independent dataset. By contrast, for smaller areas and more restricted environmental gradients, we would expect more linear I-splines, i.e. constant changes in species composition along the environmental gradients. In this case, since SCAM is more versatile than GLM, SCAM-based MS-GDM may perform better. Other parameters may also influence the results, such as the number of knots and I-splines, the objective function used in the SCAM fitting procedure, the selection of predictors, the transformation of zeta values to rescale them between 0 and 1, or the way sites are combined to compute the

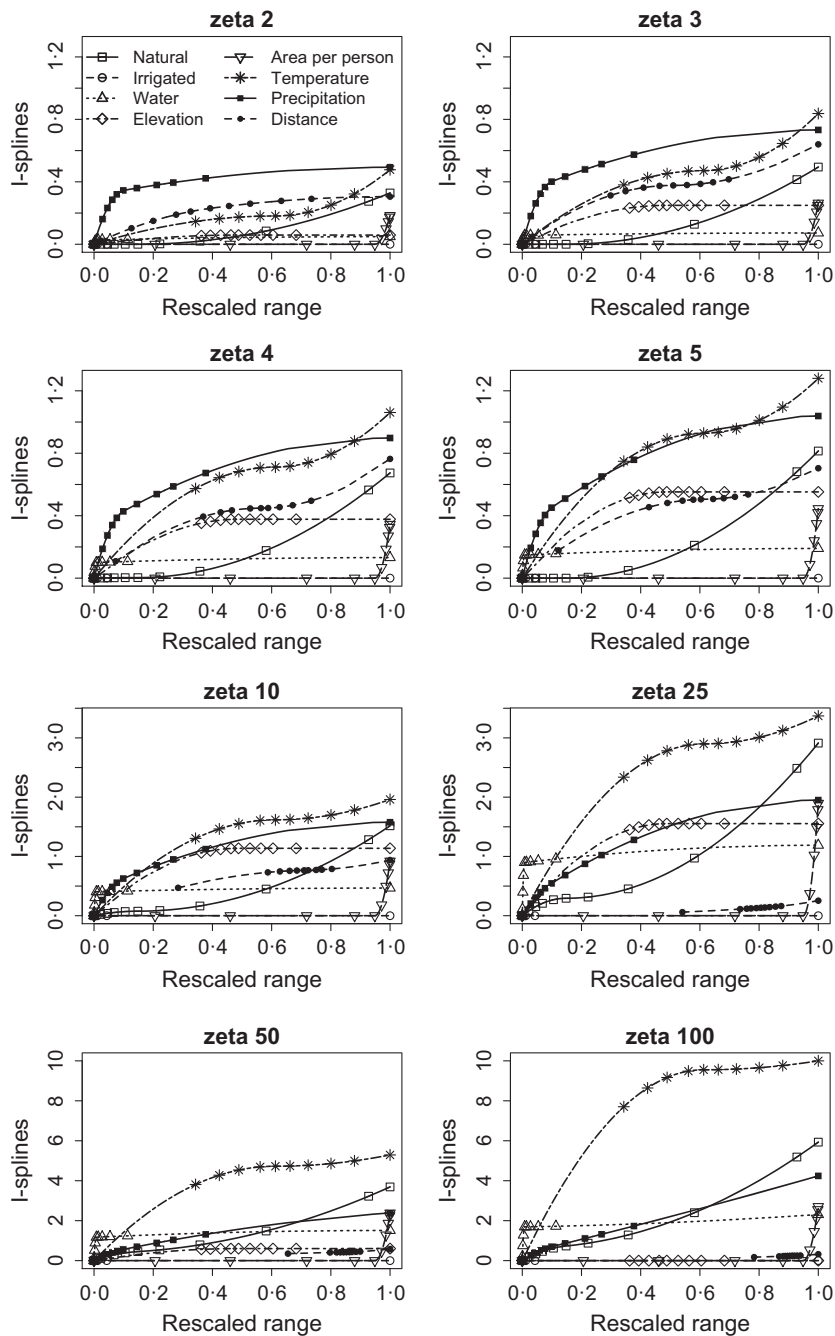


Fig. 5. I-splines showing the contribution of the seven environmental variables and distance to explaining zeta diversity, for ζ_2 – ζ_{100} , at the 100 km grain over Australia (I-splines obtained from eqn 3 after fitting eqn 5). All predictors are scaled between 0 and 1 for visualisation (0, lowest value of predictor; 1, highest value of predictor). See Tables D5 and D6 in Appendix S4 for the scaling factors.

zeta values. Exploring these different aspects of MS-GDM is of interest, but beyond the scope of this study.

The main difference between MS-GDM for low vs. higher orders of zeta was the relative importance of temperature and precipitation for species turnover. Precipitation explained most of the turnover for low orders of zeta (Fig. 5; Figs D9–D11, Appendix S4), i.e. when all species, from rare to widespread contribute to the turnover measure. By contrast, temperature became the main driver of turnover for intermediate and high orders of zeta, i.e. turnover in widespread species. While this requires more in depth ecological analysis, a speculative explanation is that, by affecting the availability of food sources such as seeds, precipitation may limit species distributions (e.g.,

Dunning & Brown 1982), and rare species may as a result be more critically related to this driver. By contrast, widespread species may be able to respond to rainfall heterogeneity via a spatial storage effect (Shoemaker & Melbourne 2016), especially in areas containing water bodies, as suggested by the fact that water bodies were more important in explaining high than low orders of zeta. Wet areas, such as floodplains, have been shown to serve as drought refugia for birds (Selwood *et al.* 2015). As a result, distributions of widespread bird species may be more related to physiological drivers such as temperature.

As intended, combining MS-GDM with the modelling of species richness (alpha diversity), provides a more complete picture of biodiversity (Socolar *et al.* 2016). Precipitation and

temperature were good predictors of both species richness and species turnover across several orders of zeta. However, richness increased linearly with temperature (Fig. 2), whereas I-spline-based MS-GDM showed that differences in temperature between warm areas were very important for species turnover, especially for low orders of zeta (Fig. 5). Richness also increased linearly with elevation, although elevation difference drove turnover only at low elevations. Finally, although the natural habitat area was only very slightly correlated with species richness at the 100 km grain over Australia, the I-splines suggest that this variable is important for species turnover for intermediate and high orders of zeta. The nonlinear shape of the splines also suggests that natural habitat area becomes increasingly important the greater the amount of natural habitat available.

One application of GDM is the prediction of species composition in unsampled sites by combining predictions of species richness and compositional dissimilarity between surveyed and unsurveyed sites estimated with GDM into an optimisation algorithm called dynamicFOAM (Mokany *et al.* 2011). Because it uses only beta diversity, this algorithm may potentially overemphasise the importance of the occurrence of rare species in estimating species composition. Considering various orders of zeta by incorporating MS-GDM into this procedure may therefore improve the overall results, by improving the prediction of widespread species shared by multiple assemblages, and have positive impacts on survey gap analyses.

Conclusion

Here, we introduce MS-GDM, which extends GDM (Ferrier *et al.* 2007) to incorporate zeta diversity as a measure of compositional similarity that is not limited to pairwise site comparison. Using measures of diversity beyond alpha and beta diversity, MS-GDM provides a more comprehensive description of the correlates of biodiversity. In particular, it captures the responses of widespread species as distinct from the response of the whole community, with the latter disproportionately influenced by the rare component of the community across site pairs in traditional measures of beta diversity. Two regression techniques were used for MS-GDM: SCAM and I-splines. Although I-spline-based MS-GDM explained the data better than SCAM-based MS-GDM, the opposite may be true for smaller study areas and environmental gradients, along which the rate of compositional turnover would remain constant. Applying MS-GDM to different orders of zeta revealed temperature as the main predictor of compositional turnover across multiple sites, i.e. for more widespread bird species, whereas differences in precipitation were more important across site pairs, i.e. for turnover driven by rare or narrow range species. Such findings may have implications for conservation planning, for predicting the impacts of climate change on biodiversity, and for understanding the consequences of turnover for ecosystem function (Schwartz *et al.* 2000). MS-GDM provides a useful addition to the biodiversity quantification and estimation toolkit, and may be

valuable for other applications of GDM, such as survey gap analysis and prediction of species composition in unsampled sites.

Authors' contributions

G.L., C.H. and M.A.M. jointly conceived the research; G.L. developed the R package, performed the analyses and drafted the manuscript; C.H. and M.A.M. contributed to the design of the experiments, the interpretation of the results, and to the manuscript preparation. All authors gave final approval for publication.

Acknowledgements

We thank Jane Elith for discussion and for comments on an earlier manuscript draft, Grant Duffy and David Baker for commenting on the paper pre-submission, and two anonymous reviewers for constructive comments. We acknowledge the BirdLife Australia Atlas Database as the data source. This work was supported by a grant to M.M. and C.H. (Australian Research Council Discovery Project DP150103017). C.H. also acknowledges support from the National Research Foundation of South Africa (nos. 89967 and 76912).

Data accessibility

The data used in the analyses is available from BirdLife Australia's Atlas of Australian Birds (Barrett *et al.* 2003; <http://birddata.birdlife.org.au/>). All relevant numerical outputs of the analyses are included in the Supporting Information.

References

- Ashcroft, M.B., Gollan, J.R., Faith, D.P., Carter, G.A., Lassau, S.A., Ginn, S.G., Bulbert, M.W. & Cassis, G. (2010) Using generalised dissimilarity models and many small samples to improve the efficiency of regional and landscape scale invertebrate sampling. *Ecological Informatics*, **5**, 124–132.
- Barrett, G., Silcocks, A., Barry, S., Cunningham, R. & Poulter, R. (2003) *The New Atlas of Australian Birds*. Royal Australasian Ornithologists Union, Melbourne.
- Baselga, A. (2013) Multiple site dissimilarity quantifies compositional heterogeneity among several sites, while average pairwise dissimilarity may be misleading. *Ecography*, **36**, 124–128.
- Calosi, P., Bilton, D.T. & Spicer, J.I. (2008) Thermal tolerance, acclimatory capacity and vulnerability to global climate change. *Biology Letters*, **4**, 99–102.
- Chao, A., Chiu, C.-H. & Hsieh, T. (2012) Proposing a resolution to debates on diversity partitioning. *Ecology*, **93**, 2037–2051.
- Chown, S.L., van Rensburg, B.J., Gaston, K.J., Rodrigues, A.S. & van Jaarsveld, A.S. (2003) Energy, species richness, and human population size: conservation implications at a national scale. *Ecological Applications*, **13**, 1233–1241.
- Clarke, K. & Gorley, R. (2015) *PERMANOVA+ for PRIMER v7: User Manual/Tutorial*. PRIMER-E, Plymouth, UK.
- Cody, M.L. (1981) Habitat selection in birds: the roles of vegetation structure, competitors, and productivity. *BioScience*, **31**, 107–113.
- De'Ath, G. (2012) The multinomial diversity model: linking Shannon diversity to multiple predictors. *Ecology*, **93**, 2286–2296.
- Dirnböck, T., Hobbs, R.J., Lambeck, R.J. & Caccetta, P. (2002) Vegetation distribution in relation to topographically driven processes in southwestern Australia. *Applied Vegetation Science*, **5**, 147–158.
- Diserud, O.H. & Ødegaard, F. (2007) A multiple-site similarity measure. *Biology Letters*, **3**, 20–22.
- Dunning, J.B. Jr & Brown, J.H. (1982) Summer rainfall and winter sparrow densities: a test of the food limitation hypothesis. *The Auk*, **99**, 123–129.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677.
- Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Faith, D.P., Minchin, P.R. & Belbin, L. (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**, 57–68.
- Ferger, S.W., Schleuning, M., Hemp, A., Howell, K.M. & Böhning-Gaese, K. (2014) Food resources and vegetation structure mediate climatic effects on species richness of birds. *Global Ecology and Biogeography*, **23**, 541–549.

- Ferrier, S. (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology*, **51**, 331–363.
- Ferrier, S., Harwood, T. & Williams, K. (2010) Using generalised dissimilarity modelling to assess potential impacts of climate change on biodiversity composition in Australia, and on the representativeness of the National Reserve System. A report to the Department of Environment, Water, Heritage and the Arts, Canberra, Australia.
- Ferrier, S., Powell, G.V., Richardson, K.S. *et al.* (2004) Mapping more of terrestrial biodiversity for global conservation assessment. *BioScience*, **54**, 1101–1109.
- Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264.
- Funk, V., Richardson, K.S. & Ferrier, S. (2005) Survey-gap analysis in expeditionary research: where do we go from here? *Biological Journal of the Linnean Society*, **85**, 549–567.
- Gaston, K.J. (2010) Valuing common species. *Science*, **327**, 154–155.
- Gaston, K.J. & Evans, K.L. (2004) Birds and people in Europe. *Proceedings of the Royal Society of London B: Biological Sciences*, **271**, 1649–1655.
- Hastie, T. & Tibshirani, R. (1990) *Generalized Additive Models*. Chapman & Hall, London, UK.
- Hawkins, B.A., Field, R., Cornell, H.V. *et al.* (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.
- Heino, J., Grönroos, M., Ilmonen, J., Karhu, T., Niva, M. & Paasivirta, L. (2013) Environmental heterogeneity and β diversity of stream macroinvertebrate communities at intermediate spatial scales. *Freshwater Science*, **32**, 142–154.
- Hortal, J., Carrascal, L.M., Triantis, K.A., Thébault, E., Meiri, S. & Sfenthourakis, S. (2013) Species richness can decrease with altitude but not with habitat diversity. *Proceedings of the National Academy of Sciences United States of America*, **110**, E2149–E2150.
- Hui, C. & McGeoch, M.A. (2014) Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. *The American Naturalist*, **184**, 684–694.
- Jetz, W., Wilcove, D.S. & Dobson, A.P. (2007) Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biology*, **5**, e157.
- Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427–2439.
- Jost, L., Chao, A. & Chazdon, R. (2011) Compositional similarity and beta diversity. *Biological Diversity: Frontiers in Measurement and Assessment* (eds A. Magurran & B. McGill), pp. 66–84. Oxford University Press, New York.
- Latombe, G., McGeoch, M.A., Nipperess, D.A. & Hui, C. (2016) zetadiv: Functions to Compute Compositional Turnover Using Zeta Diversity. R package version 1.0. Available at: <https://cran.r-project.org/package=zetadiv> (accessed 9 December 2016).
- Leitão, P.J., Schwieder, M., Suess, S. *et al.* (2015) Mapping beta diversity from space: Sparse Generalised Dissimilarity Modelling (SGDM) for analysing high-dimensional data. *Methods in Ecology and Evolution*, **6**, 764–771.
- Manion, G., Lisk, M., Ferrier, S., Nieto-Lugilde, D. & Fitzpatrick, M.C. (2016) gdm: Functions for Generalized Dissimilarity Modeling. R package version 1.2.3. Available at: <https://cran.r-project.org/package=gdm> (accessed 1 September 2016).
- Marschner, I. (2014) glm2: Fitting Generalized Linear Models. R package version 1.1.2. Available at: <https://cran.r-project.org/package=glm2> (accessed 7 October 2016).
- McGeoch, M.A. & Latombe, G. (2016) Characterizing common and range expanding species. *Journal of Biogeography*, **43**, 217–228.
- Mokany, K., Harwood, T.D., Overton, J.M., Barker, G.M. & Ferrier, S. (2011) Combining α - and β -diversity models to fill gaps in our knowledge of biodiversity. *Ecology Letters*, **14**, 1043–1051.
- Mullen, K.M. & van Stokkum, I.H.M. (2012) nnls: The Lawson-Hanson algorithm for non-negative least squares (NNLS). R package version 1.4. Available at: <https://cran.r-project.org/package=nnls> (accessed 24 March 2016).
- Nekola, J.C. & White, P.S. (1999) The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**, 867–878.
- Oksanen, J., Blanchet, F.G., Friendly, M. *et al.* (2016) vegan: Community Ecology Package. R package version 2.4-0. Available at: <https://cran.r-project.org/package=vegan> (accessed 29 July 2016).
- Prober, S.M., Hilbert, D., Ferrier, S., Dunlop, M. & Gobbett, D. (2012) Combining community-level spatial modelling and expert knowledge to inform climate adaptation in temperate grassy eucalypt woodlands and related grasslands. *Biodiversity and Conservation*, **21**, 1627–1650.
- Pya, N. (2014) scam: Shape constrained additive models. R package version 1.1-8. Available at: <https://cran.r-project.org/package=scam> (accessed 3 August 2016).
- Pya, N. & Wood, S.N. (2015) Shape constrained additive models. *Statistics and Computing*, **25**, 543–559.
- R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rahbek, C. (1997) The relationship among area, elevation, and regional species richness in neotropical birds. *The American Naturalist*, **149**, 875–902.
- Ramsay, J.O. (1988) Monotone regression splines in action. *Statistical Science*, **3**, 425–441.
- Remonti, L., Balestrieri, A. & Prigioni, C. (2009) Altitudinal gradient of Eurasian otter (*Lutra lutra*) food niche in Mediterranean habitats. *Canadian Journal of Zoology*, **87**, 285–291.
- Ricotta, C. & Pavoine, S. (2015) A multiple-site dissimilarity measure for species presence/absence data and its relationship with nestedness and turnover. *Ecological Indicators*, **54**, 203–206.
- Rose, P., Kennard, M., Sheldon, F., Moffatt, D. & Butler, G. (2016) A data-driven method for selecting candidate reference sites for stream bioassessment programs using generalised dissimilarity models. *Marine and Freshwater Research*, **67**, 440–454.
- Schwartz, M.W., Brigham, C., Hoeksema, J., Lyons, K., Mills, M. & Van Mantgem, P. (2000) Linking biodiversity to ecosystem function: implications for conservation ecology. *Oecologia*, **122**, 297–305.
- Selwood, K.E., Thomson, J.R., Clarke, R.H., McGeoch, M.A. & Mac Nally, R. (2015) Resistance and resilience of terrestrial birds in drying climates: do floodplains provide drought refugia? *Global Ecology and Biogeography*, **24**, 838–848.
- Shoemaker, L.G. & Melbourne, B.A. (2016) Linking metacommunity paradigms to spatial coexistence mechanisms. *Ecology*, **97**, 2436–2446.
- Socolar, J.B., Gilroy, J.J., Kunin, W.E. & Edwards, D.P. (2016) How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, **31**, 67–80.
- Thomassen, H.A., Cheviron, Z.A., Freedman, A.H., Harrigan, R.J., Wayne, R.K. & Smith, T.B. (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–3548.
- Warton, D.I., Wright, S.T. & Wang, Y. (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, **3**, 89–101.
- Wood, S. (2006) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton, FL, USA.
- Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, **73**, 3–36.
- Woolley, S.N., Foster, S.D., O'Hara, T.D., Wintle, B.A. & Dunstan, P.K. (2016) Characterising uncertainty in Generalised Dissimilarity Models. *Methods in Ecology and Evolution*, doi:10.1111/2041-210X.12710.

Received 13 November 2016; accepted 30 January 2017

Handling Editor: David Warton

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Data treatment.

Appendix S2. Maps and distributions of environmental variables and distance.

Appendix S3. Comparison between GDM and MS-GDM.

Appendix S4. GAM and MS-GDM for the SCAMs and I-splines, for the 100 km grain over South-East Australia and for the 25 km grain over the Australian continent and South-East Australia.