

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/230257534>

Distance-based methods for the analysis of maps produced by species distribution models

Article in *Methods in Ecology and Evolution* · June 2011

DOI: 10.1111/j.2041-210X.2011.00115.x

CITATIONS

4

READS

36

1 author:



Peter D. Wilson

Macquarie University

38 PUBLICATIONS 644 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Restore and Renew [View project](#)



Distribution of New South Wales fauna [View project](#)

All content following this page was uploaded by [Peter D. Wilson](#) on 30 September 2014.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

Distance-based methods for the analysis of maps produced by species distribution models

Peter D. Wilson*

Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia

Summary

1. Species distribution models (SDMs) are now widely applied to determine the potential distributions of species in relation to environmental covariates. Many modelling tools are available, and large sets of maps may be produced easily.
2. A wide range of methods have been developed for the comparison or analysis of raster images and SDM output maps including pairwise (two-sample) tests and overall measures of similarity such as correlation coefficients and distance measures.
3. I present an analytical framework applying a distance-based approach to the ordination and analysis of maps produced by species distribution modelling tools. The method combines aspects of image analysis with distance-based statistical tests and allows ecologists to apply familiar forms of ordination and analysis to SDM output maps. A novel method of recombining elements of information extracted from distance-based map analysis is also presented.

Key-words: analysis of variance, distance-based methods, image analysis, map comparison, principal coordinates analysis, species distribution model

Introduction

Species distribution modelling (SDM) is a popular approach to understanding the relationship between a species and its environment and for predicting changes in distribution with environmental changes. A number of highly efficient SDM tools are now available including MaxEnt (Phillips, Anderson, & Schapire 2006; Phillips & Dudik 2008) and Biomod (Thuiller *et al.* 2009) which, combined with easily accessible, large and expanding public repositories of species location data such as the Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>), make it relatively easy to produce very large sets of maps. The number of output maps produced in large-scale studies may greatly exceed human capacity for comprehending patterns of difference or similarity between sets of objects. For example, Loarie *et al.* (2008) modelled the response to climate change of 591 species of Californian plants, and Thuiller, Araújo, & Lavorel (2004) modelled the distribution of 3,377 plant and animal species.

Techniques for the statistical analysis of gridded or raster species distribution maps (including SDM outputs) have been developed on numerous occasions and span a broad range of approaches. Jardine (1972) suggested cluster analysis based on similarity or dissimilarity measures. Fewster & Buckland (2001), Ray & Burgman (2006), Hennig & Hausdorf (2006),

Warren, Glor, & Turelli (2008) and Lavigne *et al.* (2010) used measures of similarity or dissimilarity to compare maps. Correlation coefficients as a measure of similarity between pairs of maps have been used several times (Prasad, Iverson, & Liaw 2006; Termansen, McClean, & Preston 2006; Syphard & Franklin 2009). Traditional tests of statistical hypotheses have also been applied to species distribution map comparison. Levine *et al.* (2009) performed *t*-tests on mean differences in pixel values, and Syrjala (1996) compared pairs of spatial distributions using an adaptation of the Cramer-von Mises test.

However, techniques for the analysis of SDM outputs may not be limited to these approaches, and there are a number of disciplines from which methods might be drawn. This is made possible because SDM output maps, like raster maps in other spatial sciences, are equivalent to digital images where map grid cells correspond to image pixels. Physical and human geography, landscape ecology, remote sensing (Buiten 1988; Illian *et al.* 2006) and medical and general image analysis (Androustos, Plataniotis, & Venetsanopoulos 1998; Pluim, Maintz, & Viergever 2003; Howarth & Rüger 2005), for example, have faced similar challenges in the analysis of large sets of maps or images.

Borrowing from the literature on image analysis, there are three broad approaches to measuring the difference, or conversely the similarity, between pairs of SDM maps. The first measures the overall difference of binary images (Russ 1995), which requires the application of a thresholding method to convert SDM output maps to a binary (presence/absence)

*Correspondence author. E-mail: peterdonaldwilson@gmail.com
Correspondence site: <http://www.respond2articles.com/MEE/>

form. This may be carried out using a wide array of thresholding approaches (Liu *et al.*, 2005; Jiménez-Valverde & Lobo, 2007; Allouche *et al.*, 2006; Freeman & Moisen, 2008), but all thresholding methods generate false positive values in some percentage of grid cells (La Sorte & Hawkins 2007). Pairs of resulting binary-valued maps may be compared by a range of methods (Rommel & Csillag 2006; Visser & de Nijs 2006) but must inevitably suffer from some degree of confounding as a result of loss of spatial information that arises from the false positive artefact. In addition, the results of map comparison are critically dependent on the choice of threshold and threshold method.

A second approach to image comparison is to measure the difference between pairs of histograms of pixel values (Chan *et al.* 2003; Rubner *et al.*, 2000; Angulo & Serra, 2002). This avoids the issues introduced by thresholding continuously valued maps but always involves the loss of spatial information. Two histograms may be very similar in shape but represent highly divergent spatial patterns in the original maps. In addition, the choice of bin width for histograms affects histogram shape, and several approaches to selecting optimal bin widths have been proposed (Scott 1979; Wand 1997). For practical reasons, all maps in a given study converted to histogram representations must be computed using the same bin widths. Histogram-based distances, and subsequent distance-based analyses, will to some degree be dependent on the choice of bin width.

The third approach is based on pixel-by-pixel comparisons to compute a spatially explicit difference index between pairs of images (Di Gesù & Starovoitov 1999; Li & Lu 2009). The difference measure may be selected from a wide array of measures used to compare images including the Hausdorff measure (Huttenlocher, Klanderman, & Rucklidge 1993; Mémoli 2008), Euclidean distance (Wang, Zhang, & Feng 2005; Li & Lu 2009) and measures of difference related to entropy and mutual information such as Kullback–Leibler divergence and Tsallis entropy and divergence (Pluim, Maintz, & Viergever 2003; Martin *et al.* 2004; Rommel & Csillag 2006; Mohamed & Ben Hamza 2010). Hellinger distance applied to SDM output or abundance density maps (e.g. Warren, Glor, & Turelli (2008), Lavigne *et al.* (2010)) falls within this category.

An important aspect of using distance or dissimilarity measures is choosing a measure with properties suited to the question at hand. Guidance in the ecological application of similarity and dissimilarity or distance measures is provided by Gower & Legendre (1986) and Legendre & Legendre (1998). In the context of image analysis, the application of similarity and dissimilarity measures is more complex and diverse. However, some guidance on general approaches in this field is provided by Di Gesù & Starovoitov (1999) and Askoy & Haralick (2000).

A matrix of differences between all possible pairwise comparisons can be produced with the second and third approaches. The statistical analysis of distance matrices is now well advanced and includes methods for ordination, hypothesis testing and linear modelling (Rao 1995; Gower & Krzanowski 1999; Anderson & Robinson 2001, 2003; Anderson & ter

Braak 2003). In this paper, I examine the usefulness of distance-based analysis using the third approach (spatially explicit pixel-by-pixel measures) as it avoids the loss of spatial information inherent in the comparison of histograms. The performance of several measures is tested using synthetic maps, distance-based ordination and analysis of variance. The best-performing measures are then applied to two example SDM studies. I sought to answer two questions. First, how do different distance measures perform when applied to SDM map comparisons? Second, can distance-based analysis and ordination methods assist in the identification of patterns of change in large sets of SDM maps?

Methods

DIFFERENCE OR DISTANCE MEASURES

I chose to examine the properties and performance of four measures (Hellinger distance, Kullback–Leibler divergence, Tsallis divergence and Euclidean distance) because they have been widely used in images analysis and have been found to perform well in a range of map and image comparison applications. A very commonly used measure in image analysis, the Hausdorff measure, was not included because it is calculated by considering only the closest points between two objects (referred to as ‘features’ in image analysis). Therefore, this measure would not *a priori* capture details of spatial distributions in raster images without some form of preprocessing or image segmentation.

The i th map M_i in a set of maps produced by SDM software is a grid or matrix of real values with R rows indexing northing or latitude values, and C columns indexing easting or longitude values. In all that follows, it is assumed that the maps being analysed are based on the same grid system. That is, the coordinate origin and grid cell size of all maps in a set to be analysed are identical. Each of the measures described below will be applied as pixel-by-pixel comparisons of pairs of maps by indexing over the R rows and C columns of pairs of maps. The distance between two matrix representations of maps is a real number generated by a function D^k , where the superscript k indicates the distance measure defined below. In general terms, the difference between two maps M_i and M_j is

$$d_{ij}^k = D^k(M_i, M_j)$$

Some distance measures assume that the images represent bivariate probability mass functions. This requires normalization of the matrix so that the elements sum to 1, and for a map M_i is achieved by calculating:

$$S_i = \sum_{r=1}^R \sum_{c=1}^C M_i[r, c]$$

I refer to this value as the ‘intensity’ of the map by analogy with the intensity of spatial point patterns (Stoyan, Kendall, & Mecke 1995; Diggle 2003). It is also a measure of the brightness of the monochrome image interpretation of an SDM output map (see Russ 1995 for a discussion of measures of image brightness; Woods & Gonzalez 1981). Intensity is an index of the general suitability of environmental conditions for a species within the map extent. Normalization of each map by its intensity will effectively remove difference in overall habitat suitability from a distance measure and emphasize difference because of spatial pattern of normalized grid cell values. Lavigne *et al.* (2010) used similar reasoning for normalizing species abundance maps. Normalization is analogous to the removal of size to

focus on shape differences in geometric morphometrics (Small 1996; Dryden & Mardia 1998) or the removal of species richness from species diversity indices to focus on evenness (Smith & Wilson 1996). Its use in the present context does not add information but allows the information content of an SDM output map to be partitioned into meaningful components.

Hellinger distance

The Hellinger distance measures the difference between two probability density functions for continuous variables and probability mass functions for discrete variables. Hellinger distance has been used to compare image histograms (e.g. Shutin & Zlobinskaya 2010), while the application of Hellinger distance to general ecological analyses is described by Legendre & Legendre (1998). It has also been applied to map comparison by Lavigne *et al.* (2010) and Warren, Glor, & Turelli (2008). Adapting this measure for map comparison requires normalization by intensity so that the map has the properties of a bivariate probability distribution. Hellinger distance was applied using the following formula to compute a distance between maps i and j of a set:

$$d_{ij}^H = \left[0.5 * \sum_{r=1}^R \sum_{c=1}^C \left[\sqrt{\frac{M_i[r, c]}{S_i}} - \sqrt{\frac{M_j[r, c]}{S_j}} \right]^2 \right]^{1/2}$$

Kullback–Leibler divergence

Strictly referred to as a measure of divergence between two probability distributions, Kullback–Leibler divergence may be adapted to pairwise map comparison after normalization by image brightness. Kullback–Leibler divergence has been used to compare image histograms (Chan *et al.* 2003). It is also interpreted as a measure of relative entropy and is related to the idea of mutual information content for two probability distributions (Cover & Thomas 2006). The Kullback–Leibler divergence is an asymmetrical measure. That is, the difference between maps M_i and M_j will not be the same as the complementary measure between maps M_j and M_i . Common usage in information theory literature uses the asymmetrical value as a measure of ‘divergence’ (Cover & Thomas 2006), despite the fact that Kullback & Leibler (1951) used the term ‘divergence’ to refer to the symmetrical value created by adding the two complementary measures. I applied the symmetrical form to comparing two maps by the following formula, which can be easily derived from the sum of the standard formula for the complementary divergences (Cover & Thomas 2006):

$$d_{ij}^{KL} = \sum_{r=1}^R \sum_{c=1}^C \log \left(P_i[r, c] / P_j[r, c] \right) * (P_i[r, c] - P_j[r, c])$$

where the function $\log()$ refers to natural logarithms, and $P_i[r, c] = M_i[r, c] / S_i$.

Tsallis divergence

Tsallis entropy is an alternative approach to measuring relative entropy in ‘non-extensive’ systems such as images and maps (Tsallis 1999). A non-extensive physical system is one in which long-distance interactions may lead to strong autocorrelation effects, and there is some degree of self-similarity at a local scale. Tsallis entropy is a measure that accounts for the effects of non-extensivity and has been

applied in various forms to image analysis (Martin *et al.* 2004; Sun, Zhang, & Guo 2006). I calculated Tsallis divergence between pairs of SDM output maps by adapting the derivation due to Mohamed & Ben Hamza (2010) using the following formula:

$$d_{ij}^{TD} = \left(\frac{1}{1 - \alpha} \right) \left[\left(\sum_{r=1}^R \sum_{c=1}^C 0.5 (P_i[r, c] + P_j[r, c])^\alpha - 1 \right) - 0.5 \left(\left(\sum_{r=1}^R \sum_{c=1}^C P_i[r, c]^\alpha - 1 \right) + \left(\sum_{r=1}^R \sum_{c=1}^C P_j[r, c]^\alpha - 1 \right) \right) \right]$$

where $0 \leq \alpha < 1$ and $P_i[r, c] = M_i[r, c] / S_i$.

The parameter α provides a measure of the degree of non-extensivity and represents effects such as autocorrelation or self-similarity. The application of Tsallis divergence requires a value to be given for the parameter α . I chose to a value of 0.125 after preliminary trials indicated that SDM output maps consistently produced optimal values of the divergence near this value.

Euclidean distance

Pixel-wise overall Euclidean distance may be computed from the raw pixel values or on normalized values. Raw Euclidean distance for a pair of maps is expected to incorporate the influence of differences owing to brightness and differences owing to spatial distribution of brightness values. The raw distance was computed using the following formula:

$$d_{ij}^E = \left[\sum_{r=1}^R \sum_{c=1}^C (M_i(r, c) - M_j(r, c))^2 \right]^{1/2}$$

The normalized form of the Euclidean distance was calculated as follows:

$$d_{ij}^{NE} = \left[\sum_{r=1}^R \sum_{c=1}^C \left(\frac{M_i(r, c)}{S_i} - \frac{M_j(r, c)}{S_j} \right)^2 \right]^{1/2}$$

ORDINATION AND ANALYSIS METHODS

Synthetic test data

The performance of each distance measure was examined by applying each to a synthetic test data set. Three sets of synthesized test maps (50 rows by 50 columns) were generated using the *spatstat* package in R (Baddeley & Turner 2005). R-scripts to generate these data are included in Data S1 together with example realizations of each pattern as images. The first set was made of five replicates of three spatially fixed patches with a bivariate normal density distribution, one centred towards the lower left corner, one with its centroid at the middle of the spatial domain and one also centred in the lower left corner, but with its centroid slightly displaced from the first lower left patch. Within each replicate, the three patches were produced with a sequence of fixed low, medium and high maximum intensity values. An interpretable ordination of these data was expected to closely group patterns one and three (i.e. indicate high similarity) but show that both differed strongly from pattern two.

The second set of test maps was developed to examine the way in which change in location of spatial patch centroids was

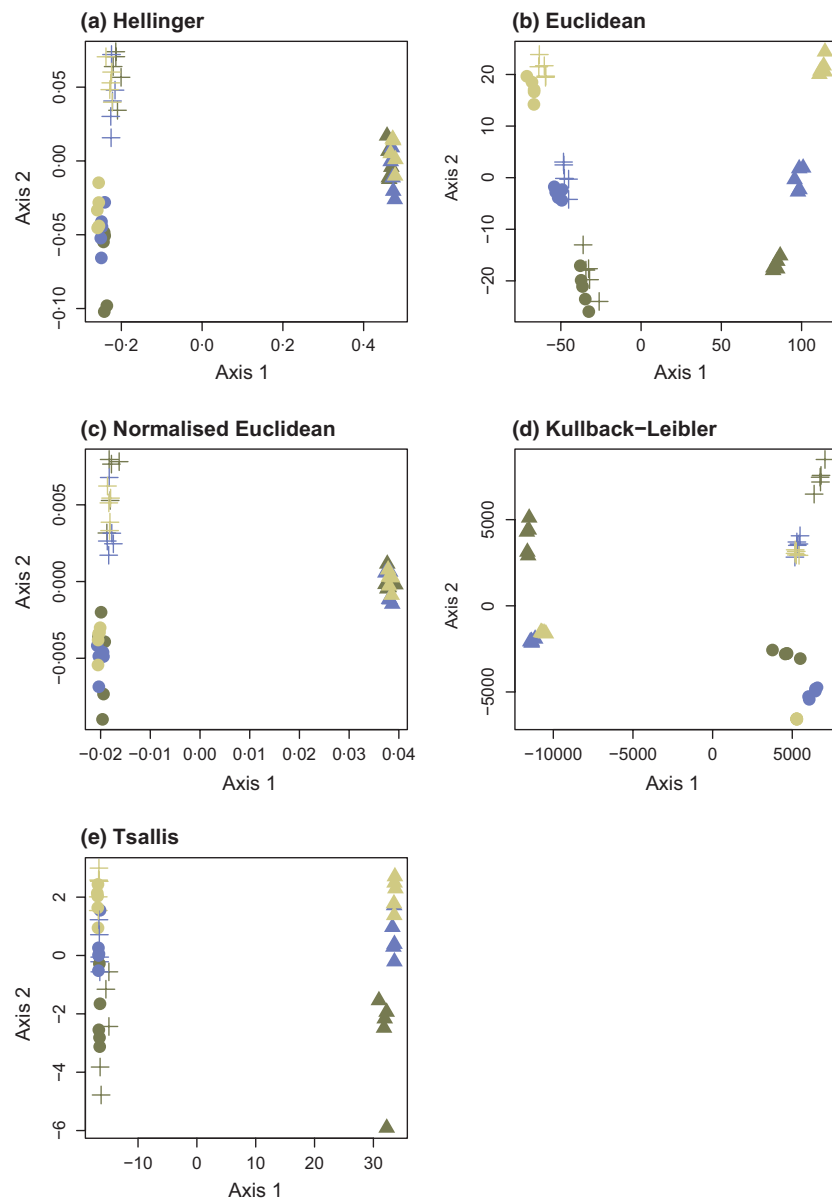


Fig. 1. Behaviour of each tested distance measure to three fixed spatial patterns with three levels of intensity represented by principal coordinate plots. Patterns 1 (●) and 3 (+) are very close in geographical space, and both differ in location from the centre of Pattern 2 (▲). Colours represent three intensity levels: Dark tan = low intensity, blue = medium intensity and light tan = high intensity. (a) Hellinger distance, (b) Euclidean distance, (c) normalized Euclidean distance, (d) Kullback–Leibler divergence and (e) Tsallis divergence. Note that the scales on each plot differ widely because each distance measure has a different range. The focus is on how accurately each portrays the expected relationship between patterns.

represented by each dissimilarity measure. Five replicates of three sequences of moving patches all with the same bivariate normal density distribution were generated. The starting point of sequences one and three corresponded to the centroids of fixed patterns one and three, and both moved in parallel a short distance apart from the lower left to the upper right of the spatial domain in two steps. The second sequence began in the lower left corner at the first track's starting point, stepped to the middle of the domain and finished in the upper left corner. An example plot of one run of the script is shown in Data S1. An interpretable ordination would be expected to show all sequences starting at the same point in the ordination plot, clustering closely at the mid-point. Sequences one and three should end very

close together, but the end point of sequence two should be distinctly different to sequences one and three.

The third set replicated the moving patch approach but varied the intensity of the patches like that used for the fixed pattern test. This test set was designed to examine the way each dissimilarity measure handled interaction between spatial position of patches and the intensity of patches. An ordination with little interaction between intensity and patch position should appear very similar to that produced from the second test data set.

Principal coordinate analysis (PCoA) was used to visualize patterns and relationships in the synthetic test data (Legendre & Legendre 1998; Cox & Cox 2001). PCoA can be applied directly to fully metric distances such as Euclidean and Hellinger distance and the symmetrical

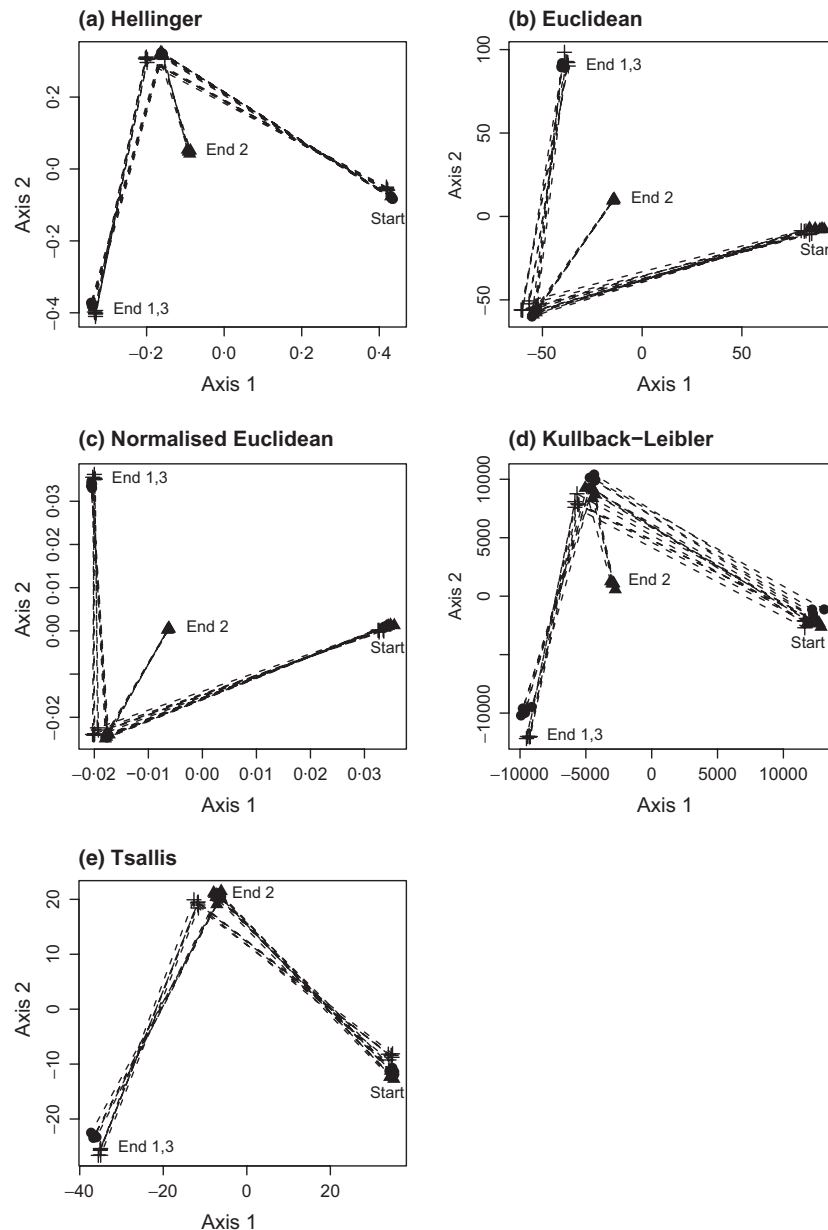


Fig. 2. Principal coordinate plots illustrating the behaviour of each tested distance measure to the same three moving spatial patterns (sequences). Intensity of moving patches was held constant at three stations along each sequence. All three sequences began very close together. Sequences 1 (●) and 3 (+) had strongly similar movement patterns within the spatial domain. Both differed sharply from sequence 2 (▲), which follows sequences 1 and 2 at steps 1 and 2 but ends at point a considerable distance from them. (a) Hellinger distance, (b) Euclidean distance, (c) normalized Euclidean distance, (d) Kullback–Leibler divergence and (e) Tsallis entropy. As noted in Fig. 1, scales on each plot differ widely because each distance measure has a different range.

form of the Kullback–Leibler divergence as they fulfil the triangle inequality condition. The triangle inequality condition requires that for three objects A, B and C, $d_{AC} \leq d_{AB} + d_{BC}$, which ensures that mathematical transformations such as PCoA preserve that relationship. PCoA may also be applied to semi-metric and general dissimilarity measures after the application of scaling adjustments to approximately fulfil the triangle inequality condition (Gower & Legendre 1986; Legendre & Legendre 1998; Legendre & Anderson 1999). I computed PCoA ordinations for each of the test and example data sets described below using the *pco* function in the *labdsv* package running in the *R* statistical environment (R Development Core Team, 2010). The *metrify* function in *labdsv* was used to trans-

form nonmetric measures to the closest equivalent Euclidean representation.

EXAMPLE APPLICATIONS

Two example applications of the methods presented in this study were made using Australian species of the tree fern genus *Cyathea* examining differences because of modelling methods, and Australian representatives of the rodent genus *Melomys*, examining differences in SDM outputs over time. In both examples, PCoA ordination was used to visualize relationships. Distance-based tests of hypotheses were made on distance or adjusted semi-metric distances (Gower &

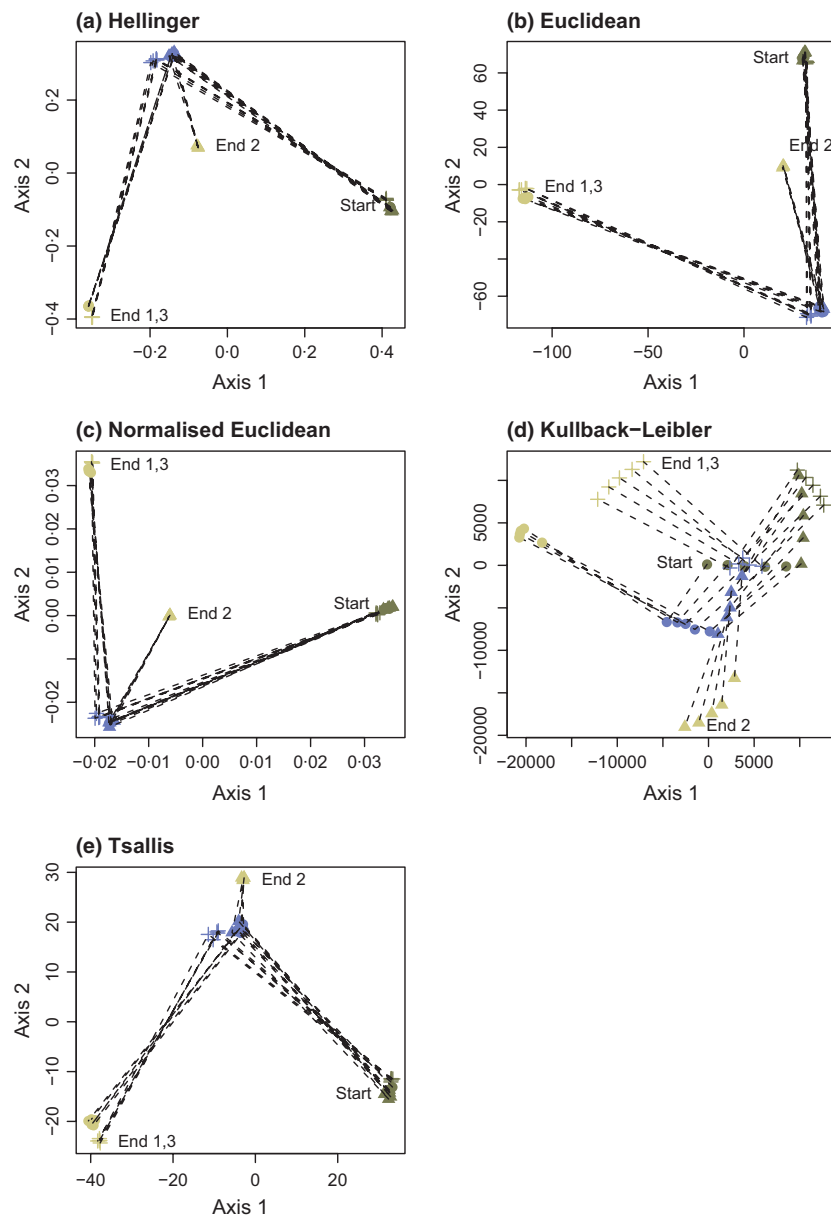


Fig. 3. Principal coordinate plots illustrating the behaviour of each tested distance measure to the same three moving spatial patterns (sequences) defined in Fig. 2, but with intensity varying systematically at three stations along each sequence. Sequences 1 (●) and 3 (+) follow each other very closely in space, and both differ sharply from Pattern 3 (▲). Colours represent three intensity levels: Dark tan = low intensity, blue = medium intensity and light tan = high intensity. (a) Hellinger distance, (b) Euclidean distance, (c) normalized Euclidean distance, (d) Kullback–Leibler divergence and (e) Tsallis entropy. As noted in Fig. 1, scales on each plot differ widely because each distance measure has a different range.

Krzanowski 1999; Legendre & Anderson 1999) using PERMANOVA (Anderson & Robinson 2001; Anderson & ter Braak 2003) available at <http://www.stat.auckland.ac.nz/~mja/Programs.htm> (last accessed 3 June 2010).

Australian species of tree ferns (genus *Cyathea*)

Eleven species of the tree fern genus *Cyathea* have been recorded from Australia, including eight endemic species (Bostock 1998; Smith *et al.* 2006): *Cyathea australis*, *C. baileyana*, *C. celebica*, *C. cooperi*, *C. cunninghamii*, *C. exilis*, *C. felix*, *C. leichhardtiana*, *C. rebecca*, *C. robertsiana* and *C. woollsiana*. Species occurrence records were retrieved from online public data portals: Australia's

Virtual Herbarium (AVH, <http://www.ersa.edu.au/avh>) and the GBIF (<http://www.gbif.org>). Ten species were modelled because *C. exilis* had only one occurrence record within Australia, and all other species had 15 or more occurrence records when duplicates were removed. Two modelling methods were used to predict the distribution of favourable climate: MaxEnt (Phillips, Anderson, & Schapire 2006; Phillips & Dudik 2008), and boosted regression trees, BRT (De'Ath 2007; Elith, Leathwick, & Hastie 2008) both of which were applied using 10 000 randomly selected background points. MaxEnt can produce indicative models with 15 records (Hernandez *et al.* 2006; Wisz *et al.*, 2008). These methods were chosen purely as examples of SDM methods that have been found to perform well, and any ensemble of methods from the many now

available could have been used as example methods (see Elith *et al.* (2006) and Franklin (2009) for overviews).

Climate data was provided by the WorldClim data set (Hijmans *et al.* 2005) for current climate conditions (i.e. climate averages over the period 1960–2000) at a grid cell size of 5 arc minutes (approximately 8 km by 8 km at -30° latitude). The full set of 19 bioclimate variables defined by Nix (1986) and Busby (1991) were used as predictor variables, downloaded from <http://www.worldclim.org>, and trimmed to a spatial extent of 100° to 160° longitude and -10° to -45° latitude, giving grids of 253 201 cells.

Five replicate models were produced by each method. For MaxEnt, this was performed using the in-built cross-validation feature with the default percentage data split, but for the BRT method, this was achieved by five cycles of randomly selecting 80% of the occurrence records and fitting models by adapting R-scripts provided by Elith, Leathwick, & Hastie (2008). PERMANOVA was performed to test for differences in spatial patterns between species, modelling method and for any interaction.

Australian species of murid rodents (genus *Melomys*)

Three species of the murid rodent genus *Melomys* are known from the Australian mainland and include *Melomys burtoni*, *M. capensis* and *M. cervinipes* (Zaret & Smith 1984). Occurrence data were combined from three online data portals: BioMaps (<http://www.biomaps.org>), OZCAM (<http://www.ozcam.org>) and GBIF (<http://www.gbif.org>). The spatial distribution of favourable climate conditions under current climate was modelled using the WorldClim 5 arc minute data described above. Only five replicate MaxEnt models were produced for each species using default settings except that threshold and hinge features were turned off. All 19 bioclimate variables were used as predictors. Models were projected onto downscaled future climate from the CSIRO Mark 3.5 General Circulation Model (Collier *et al.* 2008), which was used in the Inter-governmental Panel on Climate Change's Fourth Assessment Report (Solomon *et al.* 2007). Downscaling to the WorldClim 5 arc minute grid was performed using bicubic spline interpolation (Press *et al.* 2002). Fitted MaxEnt models were projected onto climate averages for the decades centred on 2020 and 2050. PERMANOVA was performed to test for differences in spatial patterns between species, times and any interaction between these.

Results

SYNTHETIC TEST DATA

Ordination of the fixed patterns (Fig. 1) demonstrated that Hellinger and normalized Euclidean distances (Fig. 1a,c) behaved as expected and were very similar in their representation of similarities and differences. Both were little affected by variation in overall intensity. Raw Euclidean distance (Fig. 1b) separated the three patterns according to expectations, but is clearly influenced by the trend in intensity. The Kullback–Leibler-based ordination (Fig. 1d) separated patterns 1 and 3 and correctly indicated the difference between patterns 1 and 3 and pattern 2. The Tsallis divergence-based ordination performed relatively poorly, clearly separating patterns 1 and 3 from pattern 2, but was unable to resolve patterns 1 and 3 (Fig. 1e). The results for spatially fixed synthetic patterns clearly indicated

that the best results were provided by ordinations based on normalized full metric distances (Hellinger and normalized Euclidean).

A similar set of results were apparent for the constant density moving pattern test (Fig. 2), although the raw Euclidean distance-based ordination performed equally as well as the two normalized full metric distances (Hellinger and normalized Euclidean). Kullback–Leibler divergence (Fig. 2d) provided clear differentiation of sequence 2 from sequences 1 and 3 and was very similar in conformation to the Hellinger and two Euclidean ordinations. Tsallis divergence (Fig. 2e) produced an ordination plot that represented sequences 1 and 3 in the same manner as the other measures, but confounded the end point for sequence 2 with the mid-point of the three sequences.

The second moving pattern test (Fig. 3) that added variation in intensity with spatial changes produced highly similar results for the Hellinger and two Euclidean distance measures. Kullback–Leibler divergence was influenced by an interaction between intensity and spatial shifts in patch centroids, but did indicate that the end point of the second sequence differed from that for sequences 1 and 3. Tsallis divergence produced an ordination that was similar to Hellinger and Euclidean ordinations, but in this instance did separate the end point for sequence 2 from the mid-point of all sequences.

In summary, the results indicated that the Hellinger and the two Euclidean distances produced stable and easily interpreted ordinations. Kullback–Leibler divergence and Tsallis divergence ordinations were not as easily interpreted and showed strong interactions between spatial pattern and variable intensity of test patches even though they were computed using normalized maps.

EXAMPLE APPLICATIONS

Cyathea species and SDM method

The ordination of Hellinger distances (Fig. 4) showed that there were similarities in maps produced by BRT and MaxEnt methods because, for each species, BRT points and MaxEnt points were associated in the plot. However, the analysis also illustrated that maps produced by each method (provided in Data S1) were never completely concordant and that distances between clusters of points for each method varied widely amongst species. Analysis of variance (Table 1) confirmed the indicated pattern seen in the PCoA plot. The presence of a strongly significant interaction between species and model suggested that one or more species 'responded' differently to each modelling method.

Melomys species and climate change

The PCoA plot for the *Melomys* data set indicated a distinct separation between the three species that was linked to the known distribution of species. *M. burtoni* is widely distributed across northern Australia and extends down the east coast to

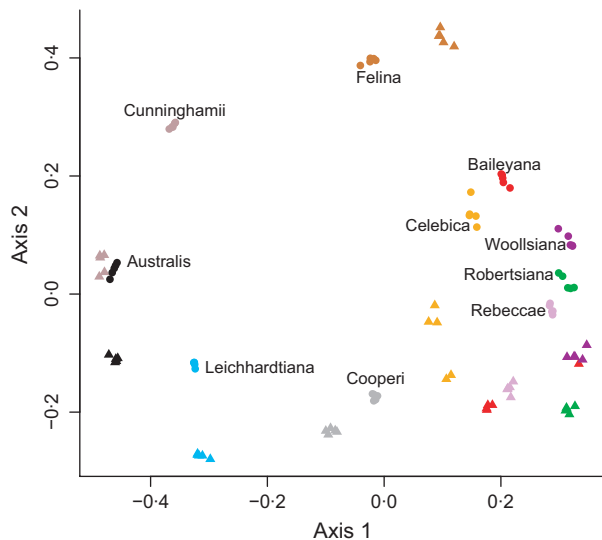


Fig. 4. Principal coordinates plot of relationships between projected distributions of 10 *Cyathea* species and species distribution model (SDM) method. Circles (●) represent five replicate boosted regression tree (BRT) models and triangles (▲) five replicate MaxEnt models for each species where colour links points for each species. The plot suggests broad similarity between modelling methods, but highlights some differences. For example, there was a large difference between BRT and MaxEnt models for *C. baileyana* and an outlying MaxEnt model for this species.

Table 1. Permutational multivariate analysis of variance using Hellinger distance for *Cyathea* SDM maps

Source	df	SS	MS	F	P(perm)	P(MC)
Model	1	2.033	2.033	297.354	0.001	0.001
Species	9	17.186	1.910	279.363	0.001	0.001
Model × Species	9	4.597	0.511	74.720	0.001	0.001
Residual	80	0.547	0.007			
Total	99	24.362				

SDM, species distribution models.

approximately Taree, New South Wales (Breed & Ford 2007; van Dyck & Strahan 2008). *M. capensis* is restricted to the northern tip of Cape York. These two very differently distributed species (reflected by the wide separation on the ordination of Fig. 5) showed very little shift in the predicted distribution of suitable future climates, which was represented in Fig. 5 by the limited movement of points along the time sequence. These two species indicated that distance between times along a sequence did not represent changes in absolute spatial extent of favourable conditions, but reflected *relative* change in their distribution (see SDM output maps in Data S1). In contrast, *M. cervinipes*, which is distributed down the east coast of Australia from Cape York to just north of Sydney, showed very large changes in spatial pattern under climate change. Inspection of SDM maps confirmed that bioclimate suitable for this species is predicted to contract progressively south with warming and drying over time. PerMANOVA results supported these interpretations (Table 2).

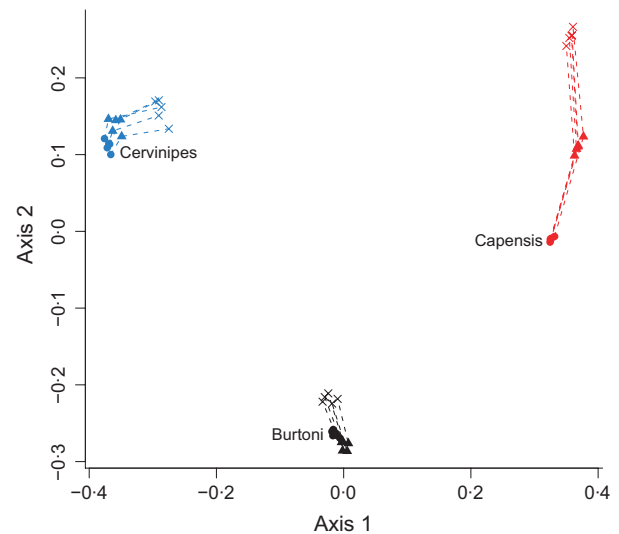


Fig. 5. Principal coordinates plot for three *Melomys* species using Hellinger distance tracking changes in spatial distribution at three climate epochs represented by five replicate MaxEnt models fitted to current climate and projected onto predicted climate for 2020 and 2050. Circles (●) indicate current climate, triangles (▲) climate in 2020 and crosses (×) climate in 2050. *M. burtoni* and *M. cervinipes* show limited change in spatial pattern under predicted climate change, but *M. capensis* undergoes large shifts in spatial pattern through time.

Discussion

The approach described in this paper advances the development of distance- and dissimilarity-based analytical methods in ecology, which has a long history (Legendre & Legendre 1998). The most recent advances in this field include the development of distance-based multivariate analysis of variance and linear modelling (Gower & Krzanowski 1999; Legendre & Anderson 1999; Anderson 2001a,b). The technique presented here is a natural extension of distance-based methods to the direct comparison of SDM output maps. Several difference measures were tested, and Hellinger distance and normalized Euclidean distance were found to be the most effective at representing spatial differences between SDM maps.

There are several advantages to the methods described in the present study. First, the basic method may be applied using any one of numerous distance or dissimilarity measures providing flexibility to the analyst. Second, a number of appropriate ordination methods may be applied to distance or dissimilarity matrices including PCoA and nonmetric multidimensional scaling (NMDS) (Legendre & Legendre 1998). Third, it allows the use of ordination and analysis methods familiar to ecologists to examine complex relationships and therefore is an advance over the two-sample methods of Levine *et al.* (2009) and Syrjala (1996), and the simple correlation approach applied by Prasad, Iverson, & Liaw (2006), Syphard & Franklin (2009) and Termansen, McClean, & Preston (2006). Fourth, the method has universal application to any set of SDM output maps produced by any modelling method. This allows threshold-free intermodel comparison of predicted spatial patterns and is an important adjunct to traditional

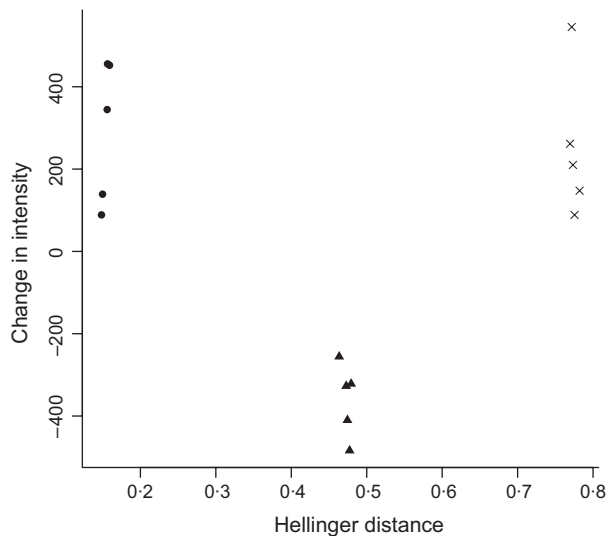


Fig. 6. An illustration of one suggested use for the intensity and distance information based on the *Melomys* results. Change in intensity between the current and 2050 maps for each species is plotted against the corresponding Hellinger distance. *M. burtoni* (●) is predicted to experience an increase in favourable conditions (positive change in intensity) but little spatial change. *M. capensis* (▲) is predicted to show moderate spatial changes but also a decrease in favourable climate conditions, while *M. cervinipes* (×) may see an increase in favourable climates but large changes in the spatial distribution of those conditions.

Table 2. Permutational multivariate analysis of variance using Hellinger distance for *Melomys* SDM maps

Source	df	SS	MS	F	P(perm)	P(MC)
Species	2	5.066	2.533	984.996	0.001	0.001
Time	2	0.392	0.196	76.153	0.001	0.001
Species × Time	4	0.713	0.178	69.353	0.001	0.001
Residual	36	0.093	0.003			
Total	44	6.263				

SDM, species distribution models.

model quality indices such as area under the receiver operating curve (AUC) (Fielding & Bell 1997). This is a distinct advantage because similarly high AUC scores, which are indicative of overall good model performance relative to the set of occurrence records, may be associated with SDM outputs with distinctly different spatial patterns. The distance-based approach presented here directly measures the similarities or differences between the spatial patterns. Finally, techniques such as PERMANOVA and related forms of distance-based linear modelling enable the design of studies that may allow the influence of known sources of bias or variability in SDM outputs to be examined directly (Heikkinen *et al.* 2006; Ray & Burgman 2006; Guisan *et al.* 2007; Beaumont, Hughes, & Pitman 2008; Graham *et al.*, 2008).

There are, however, some constraints on its application. The number of pairwise distances to be computed rises quadratically in the number of items to be compared leading to very

large matrices (i.e. $n(n-1)/2$ for symmetric measures between n maps). The potential exists for computing constraints on the size and complexity of studies. For example, computing the Hellinger distance matrix for the *Melomys* example (990 distances) took 2 min 9 s using a custom-written compiled program. The same computations for the *Cyathea* example involved the computation of 4950 unique distances and took 11 min 13 s. Using these timings indicates that, say, increasing the number of cross-validations to 10 in the *Cyathea* example would increase the number of unique distances to 39 600 and take approximately 1.5 h. Also, large collections of items to be analysed may produce complex or crowded ordination plots that can be difficult to interpret.

The method presented here extracts two types of information from an ensemble of SDM output maps – intensity and spatial difference in the form of a distance measure. Intensity measures the overall level of favourable environmental conditions (conditional on the environmental covariates used in the SDMs). A low intensity indicates a very limited presence of suitable conditions for a species, which may range between a small high-quality patch within the spatial domain of a study or everywhere may be of marginal quality. That is, map intensity (homologous with image brightness) represents an overall or average measure of suitability with spatial information removed. When a normalized distance measure is used to measure difference in spatial pattern, the corresponding difference in intensity, which is largely independent of the information in the distance measure, can be used in a scatterplot to provide insights into patterns of change in the distribution of environmental suitability (Fig. 6).

I have shown that a natural extension to distance-based ordination and analysis can be usefully applied to SDM output maps. The approach is intuitive (drawing on analytical concepts familiar to ecologists) and easily implemented using widely available tools. The information extracted using the method also permits informative displays of changes in components of spatial pattern. There are many areas for further investigation using the testing and evaluation framework presented here. These might include the performance of other distance or dissimilarity measures, other ordination methods, and the impact on spatial distributions of collinearity or concurrency in predictor variables.

Acknowledgements

I gratefully acknowledge [Jeremy VanDerWal](#) for discussions at a very early stage that helped focus my work. This work was supported by a postdoctoral fellowship at Macquarie University funded by an Australian Research Council Linkage Grant to Michelle Leishman, Lesley Hughes and Paul Downey with the New South Wales Department of Environment, Climate Change and Water as partner organization. [Michelle Leishman](#), Lesley Hughes and Jessica O'Donnell provided extensive comments on the manuscript. I thank [Jane Elith](#) and an anonymous reviewer for helpful criticism leading to improvements to this paper.

References

- Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

- Anderson, M.J. (2001a) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, **26**, 32–46.
- Anderson, M.J. (2001b) Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 626–639.
- Anderson, M.J. & Robinson, J. (2001) Permutation tests for linear models. *Australian and New Zealand Journal of Statistics*, **43**, 75–88.
- Anderson, M.J. & Robinson, J. (2003) Generalized discrimination analysis based on distances. *Australian and New Zealand Journal of Statistics*, **45**, 301–318.
- Anderson, M.J. & ter Braak, C.J.F. (2003) Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*, **73**, 85–113.
- Androutsos, D., Plataniotis, K.N. & Venetsanopoulos, A.N. (1998) Distance measures for colour image retrieval. *International Conference on Image Processing*, pp. 770–774. IEEE Computer Society, Chicago, IL, USA.
- Angulo, J. & Serra, J. (2002) Morphological color size distributions for image classification and retrieval. *Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems)*, Ghent, Belgium, September 9–11, 2002, pp. S00–1–S00–8, Ghent, Belgium.
- Askoy, S. & Haralick, R.M. (2000) Probabilistic vs. geometric similarity measures for image retrieval. *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 357–362. IEEE, Hilton Head Island, South Carolina, USA.
- Baddeley, A. & Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**, 1–42.
- Beaumont, L.J., Hughes, L. & Pitman, A.J. (2008) Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters*, **11**, 1135–1146.
- Bostock, P.D. (1998) Cyatheaaceae. *Flora of Australia Online*. Australian Biological Resources Study. Last accessed 4 June 2010.
- Breed, B. & Ford, F. (2007) *Native Mice and Rats*, 1st edn. CSIRO Publishing, Collingwood, Victoria.
- Buiten, H.J. (1988) Matching and mapping of remote sensing images: aspects of methodology and quality. *Proceedings 16th ISPRS Congress*, pp. 321–330. Kyoto, Japan.
- Busby, J.R. (1991) BIOCLIM – a bioclimate analysis and prediction system. *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (ed. by C.R. Margules & M.P. Austin), pp. 64–68. CSIRO, Canberra.
- Chan, H.-M., Chung, A.C.S., Yu, S.C.H., Norbash, A. & Wells III, W.M. (2003) Multi-modal image registration by minimizing Kullback–Leibler distance between expected and observed joint class histograms. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)*. IEEE Computer Society.
- Collier, M.A., Dix, M.R., Hirst, A.C., Davies, H.L., Elliott, T.I., Gordon, H.B., Morgan, P.J.L., O'Farrell, S.P., Rotsteyn, L.D. & Watterson, I.G. (2008) IPCC Standard Output from the CSIRO Mk3.0 Climate System Model. CSIRO Marine and Atmospheric Research Paper 008.
- Cover, T.M. & Thomas, J.A. (2006) *Elements of Information Theory*, 2nd edn. John Wiley and Sons, Hoboken, New Jersey.
- Cox, T.F. & Cox, M.A.A. (2001) *Multidimensional Scaling*, 2nd edn. Chapman and Hall/CRC, Boca Raton, Florida.
- De'Ath, G. (2007) Boosted regression trees for ecological modeling and prediction. *Ecology*, **88**, 243–251.
- Di Gesù, V. & Starovoitov, V. (1999) Distance-based functions for image comparison. *Pattern Recognition Letters*, **20**, 207–214.
- Diggle, P.J. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn. Arnold, London.
- Dryden, I.L. & Mardia, K.V. (1998) *Statistical Shape Analysis*, 1st edn. John Wiley & Sons, Chichester, UK.
- van Dyck, S. & Strahan, R. (2008) *The Mammals of Australia*. New Holland, Sydney.
- Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., L., J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T. & Phillips, S.J. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Fewster, R.M. & Buckland, S.T. (2001) Similarity indices for spatial ecological data. *Biometrics*, **57**, 495–501.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Franklin, J. (2009) *Mapping Species Distributions: Spatial Inference and Prediction*, 1st edn. Cambridge University Press, New York.
- Freeman, E.A. & Moisen, G.G. (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48–58.
- Gower, J.C. & Krzanowski, W.J. (1999) Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Applied Statistics*, **48**, 505–519.
- Gower, J.C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification*, **3**, 5–48.
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T. & Loiselle, B.A. & The NCEAS Predicting Species Distributions Working Group (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239–247.
- Guisan, A., Graham, C.H., Elith, J., Huettmann, F., Dudik, M., Ferrier, S., Hijmans, R.J., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B., Manion, G., Moritz, C., Nakamura, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Periera, R., Schapire, R.E., Williams, S.E., Wisz, M.S. & Zimmerman, N.E. (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, **13**, 332–340.
- Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751–777.
- Hennig, C. & Hausdorf, B. (2006) A robust distance coefficient between distribution areas incorporating geographic distances. *Systematic Biology*, **55**, 170–175.
- Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Howarth, P. & Rüger, S. (2005) Fractional distance measures for content-based image retrieval. *Lecture Notes in Computer Science*, **3408**, 447–456.
- Huttenlocher, D.P., Klanderman, G.A. & Rucklidge, W.J. (1993) Comparing images using the Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 850–863.
- Illian, J., Benson, E., Crawford, J. & Staines, H. (2006) Principal components analysis for spatial point processes – assessing the appropriateness of the approach in an ecological context. *Case Studies in Spatial Point Pattern Process Modelling* (ed. by A. Baddeley, P. Gregori, J. Mateu, R. Stoica & D. Stoyan), pp. 135–150. Springer, New York.
- Jardine, N. (1972) Computational methods in the study of plant distributions. *Taxonomy, Phytogeography and Evolution* (ed. by D.H. Valentine), pp. 381–393. Academic Press, London.
- Jiménez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, **31**, 361–369.
- Kullback, S. & Leibler, R.A. (1951) On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.
- La Sorte, F.A. & Hawkins, B.A. (2007) Range maps and species richness patterns: errors of commission and estimates of uncertainty. *Ecography*, **30**, 649–662.
- Lavigne, C., Ricci, B., Franck, P. & Senoussi, R. (2010) Spatial analyses of ecological count data: a density map comparison approach. *Basic and Applied Ecology*, **11**, 734–742.
- Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*, 2nd English edn. Elsevier, Amsterdam.
- Levine, R.S., Yorita, K.L., Walsh, M.C. & Reynolds, M.G. (2009) A method for statistically comparing spatial distribution maps. *International Journal of Health Geographics*, **8**, 7.
- Li, J. & Lu, B.-L. (2009) An adaptive image Euclidean distance. *Pattern Recognition*, **42**, 349–357.
- Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.
- Loarie, S.R., Carter, B.E., Hayhoe, K., McMahon, S., Moe, R., Knight, C.A. & Ackerly, D.D. (2008) Climate change and the future of California's endemic flora. *PLoS ONE*, **3**, e2502.
- Martin, S., Morison, G., Nailon, W. & Durrani, T. (2004) Fast and accurate image registration using Tsallis entropy and simultaneous perturbation stochastic approximation. *Electronics Letters*, **40**, 595–597.

- Mémoli, F. (2008) Gromov-Hausdorff distances in Euclidean spaces. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–8. IEEE Computer Society, Anchorage, Alaska, USA.
- Mohamed, W. & Ben Hamza, A. (2010) Medical image registration using stochastic optimization. *Optics and Lasers in Engineering*, **48**, 1213–1223.
- Nix, H.A. (1986) A biogeographic analysis of the Australian elapid snakes. *Atlas of Elapid Snakes* (ed. by R. Longmore), pp. 4–15. AGPS, Canberra.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Phillips, S.J. & Dudík, M. (2008) Modeling of species distribution with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Plum, J.P.W., Maintz, J.B.A. & Viergever, M.A. (2003) Mutual-Information-Based Registration of Medical Images: A Survey. *IEEE Transactions on Medical Imaging*, **22**, 986–1004.
- Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**, 181–199.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (2002) *Numerical Recipes in C++*. The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge, UK.
- R Development Core Team. (2010) *R: A Language and Environment for Statistical Computing*. Version 2.11.0. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C.R. (1995) A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Öttestiöl*, **19**, 23–63.
- Ray, N. & Burgman, M. (2006) Subjective uncertainties in habitat suitability maps. *Ecological Modelling*, **195**, 172–186.
- Remmel, T.K. & Csillag, F. (2006) Mutual information spectra for comparing categorical maps. *International Journal of Remote Sensing*, **27**, 1425–1452.
- Rubner, Y., Tomasi, C. & Guibas, L.J. (2000) The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision*, **40**, 99–121.
- Russ, J.C. (1995) *The Image Processing Handbook*, 2nd edn. CRC Press, Boca Raton, Florida.
- Scott, D.W. (1979) On optimal and data-based histograms. *Biometrika*, **66**, 605–610.
- Shutin, D. & Zlobinskaya, O. (2010) Application of information-theoretic measures to quantitative analysis of immunofluorescent microscope imaging. *Computer Methods and Programs in Biomedicine*, **97**, 114–129.
- Small, C.G. (1996) *The Statistical Theory of Shape*, 1st edn. Springer, New York.
- Smith, B. & Wilson, J.B. (1996) A consumer's guide to evenness indices. *Oikos*, **76**, 70–82.
- Smith, A.R., Pryer, K.M., Schuettpelz, E., Korall, P., Schneider, H. & Wolf, P.G. (2006) A classification for extant ferns. *Taxon*, **55**, 705–731.
- Solomon, S., Qin, D., Manning, M., Marquis, M., Averyt, K., Tignor, M.M.B., Miller Jr, H.L. & Chen, Z. (2007) *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, Cambridge.
- Stoyan, D., Kendall, W.S. & Mecke, J. (1995) *Stochastic Geometry and its Applications*, 2nd edn. John Wiley & Sons, Chichester, UK.
- Sun, S., Zhang, L. & Guo, C. (2006) Medical Image Registration by Minimizing Divergence Measure Based on Tsallis Entropy. *International Journal of Biological and Life Sciences*, **2**, 75–80.
- Syphard, A.D. & Franklin, J. (2009) Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography*, **32**, 907–918.
- Syrjala, S.E. (1996) A statistical test for a difference between the spatial distributions of two populations. *Ecology*, **77**, 75–80.
- Termansen, M., McClean, C.J. & Preston, C.D. (2006) The use of genetic algorithms and Bayesian classification to model species distributions. *Ecological Modelling*, **192**, 410–424.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, **31**, 353–361.
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Tsallis, C. (1999) Nonextensive statistics: theoretical, experimental and computational evidences and connections. *arXiv:cond-mat*, **9903356v1**, 1–47.
- Visser, H. & de Nijs, T. (2006) The map comparison kit. *Environmental Modelling & Software*, **21**, 346–358.
- Wand, M.P. (1997) Data-based choice of histogram bin width. *The American Statistician*, **51**, 59–64.
- Wang, L., Zhang, Y. & Feng, J. (2005) On the Euclidean distance of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 1334–1339.
- Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalence versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H. & Guisan, A. & NCEAS Predicting Species Distributions Working Group (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.
- Woods, R.E. & Gonzalez, R.C. (1981) Real-time digital image enhancement. *Proceedings of the IEEE*, **69**, 643–654.
- Zaret, T.M. & Smith, E.P. (1984) On measuring niches and not measuring them. *Evolutionary Ecology of Tropical Freshwater Fishes* (ed. by T.M. Zaret), pp. 127–137. Dr W. Junk, The Hague.

Received 28 June 2010; accepted 6 April 2011

Handling Editor: Robert Freckleton

Supporting Information

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

Data S1. R-scripts and examples of test maps.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.