



Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models

Mohsen B. Mesgaran¹, Roger D. Cousens¹ and Bruce L. Webber^{2,3*}

¹Department of Resource Management & Geography, The University of Melbourne, Melbourne, Vic. 3010, Australia, ²CSIRO Ecosystem Sciences, Wembley, WA 6913, Australia, ³School of Plant Biology, The University of Western Australia, Crawley, WA 6009, Australia

ABSTRACT

Aim Correlative species distribution models (SDMs) often involve some degree of projection into novel covariate space (i.e. extrapolation), because calibration data may not encompass the entire space of interest. Most methods for identifying extrapolation focus on the range of each model covariate individually. However, extrapolation can occur that is well within the range of univariate variation, but which exhibits novel combinations between covariates. Our objective was to develop a tool that can detect, distinguish and quantify these two types of novelties: novel univariate range and novel combinations of covariates.

Location Global, Australia, South Africa.

Methods We developed a new multivariate statistical tool, based on the Mahalanobis distance, which measures the similarity between the reference and projection domains by accounting for both the deviation from the mean and the correlation between variables. The method also provides an assessment tool for the detection of the most influential covariates leading to dissimilarity. As an example application, we modelled an Australian shrub (*Acacia cyclops*) widely introduced to other countries and compared reference data, global distribution data and both types of model extrapolation against the projection globally and in South Africa.

Results The new tool successfully detected and quantified the degree of dissimilarity for points that were either outside the univariate range or formed novel covariate combinations (correlations) but were still within the univariate range of covariates. For *A. cyclops*, more than half of the points (6617 of 10,785) from the global projection space that were found to lie within the univariate range of reference data exhibited distorted correlations. Not all the climate covariates used for modelling contributed to novelty equally over the geographical space of the model projection.

Main conclusions Identifying non-analogous environments is a critical component of model interrogation. Our extrapolation detection (ExDet) tool can be used as a quantitative method for exploring novelty and interpreting the projections from correlative SDMs and is available for free download as stand-alone software from <http://www.climond.org/exdet>.

Keywords

Correlation, ExDet, Mahalanobis distance, MAXENT, MESS map, model extrapolation, model interrogation, niche modelling, non-equilibrium settings, novel environment, SDM.

*Correspondence: Bruce L. Webber, CSIRO Ecosystem Sciences, Private Bag 5, Wembley, WA 6913, Australia.
E-mail: bruce.webber@csiro.au

INTRODUCTION

Correlative species distribution models (SDMs; sometimes also known as bioclimatic envelope models or niche models; Peterson & Soberón, 2012) are now commonly used to address a wide variety of questions, including conservation prioritization (Thuiller *et al.*, 2004a; Rhodes *et al.*, 2006), anticipation of climate change impacts on species range shifts and biodiversity (Stralberg *et al.*, 2009; Webber *et al.*, 2011), invasive species potential range expansion (Ebeling *et al.*, 2008; Václavík & Meentemeyer, 2009), epidemic risks from pathogens (Meentemeyer *et al.*, 2008) and phylogeography (Scoble & Lowe, 2010; Schorr *et al.*, 2012). Despite the acknowledged popularity and potential usefulness of SDMs (Guisan & Zimmermann, 2000; Jeschke & Strayer, 2008; Araújo & Peterson, 2012), there are several difficulties and limitations, both methodological and conceptual, with which users need to be familiar (Venette *et al.*, 2010), but that remain frequently overlooked by both modellers and end-users alike (Sutherland & Bourne, 2009; Kriticos *et al.*, 2013). A particular concern has been the extent to which an SDM is transferable to places or times different from those used for model calibration and how this limitation influences the interpretation of model results (Davis *et al.*, 1998; Araújo & Guisan, 2006; Randin *et al.*, 2006; Dormann, 2007; Peterson *et al.*, 2007; Rocchini *et al.*, 2011). Correlative SDMs often require some degree of extrapolation (i.e. projection into novel regions of covariate space; Elith & Leathwick, 2009) because a species' distribution might not be at equilibrium with the environment (e.g. as a result of recent introduction, dispersal limitations or biotic constraints) and the species' spatial occurrence data may not represent the complete distribution of species (e.g. because of poor monitoring or biased sampling; Franklin, 2010; Araújo & Peterson, 2012).

Particular applications of correlative SDMs, however, have the potential to result in significant extrapolation when the model is forced into conditions non-analogous to those under which it has been calibrated (Thuiller *et al.*, 2004b; Dormann, 2007; Williams & Jackson, 2007; Williams *et al.*, 2007; Franklin, 2010). For example, it is a common practice to train a model using environmental data from within the native range of an invasive alien species and then to project this calibrated model to the introduced region to delineate the potential region where naturalization might be more likely (e.g. Strubbe & Matthysen, 2009). Similarly, current environmental correlates of a species range can be used to forecast the potential future distribution of species in response to ongoing global climate change (e.g. Webber *et al.*, 2011). The application of an SDM beyond the niche breadth used to formulate it is very risky (Elith *et al.*, 2010) and might be ecologically and statistically invalid (Davis *et al.*, 1998; Thuiller *et al.*, 2004b; Dormann, 2007; Fitzpatrick & Hargrove, 2009). However, the requirements of the research mean that the situation often cannot be avoided (Araújo & Guisan, 2006; Elith *et al.*, 2010). It is therefore important to develop a better understanding of dragons

potentially lurking in the uncharted waters of novel covariate space.

Recognizing the non-homologous climate extrapolation issue, it has been suggested that a map indicating projected suitability should be accompanied by some measure of uncertainty: a counterpart map, for example, which indicates regions where the model has been extrapolated (Platts *et al.*, 2008; Fitzpatrick & Hargrove, 2009). The development of methods to address uncertainty has lagged behind the rate of adoption of SDMs, and some widely used software packages lack any ability to explore these concerns. Several useful methods, however, have been suggested recently to identify and visualize novel environmental conditions (Williams *et al.*, 2007; Platts *et al.*, 2008; Fitzpatrick & Hargrove, 2009; Elith *et al.*, 2010; Zurell *et al.*, 2012). A good example is the Multivariate Environmental Similarity Surface feature (or MESS map, *sensu* Elith *et al.*, 2010) implemented in recent versions of MAXENT (3.3.2 and later; Phillips *et al.*, 2006). The MESS tool calculates the similarity of each point in the region of projection to a set of reference points (e.g. background data) and maps the results (Elith *et al.*, 2010). MESS maps produced by MAXENT can help the user to identify extrapolated areas and thus provide a quantitative measure of projection uncertainty. However, the detection of extrapolation or 'dissimilar' points by the MESS tool and most others (e.g. Platts *et al.*, 2008) considers only the ranges of individual (univariate) predictors; they take no account of the correlation structure (i.e. new multivariate combinations) of the many covariates that may be included in the model.

For a univariate situation, it could be valid to look only at the range of values of any given covariate to identify unusual points. However, SDMs are calibrated within the multivariate space of often large numbers of covariates which can be highly correlated. For example, correlation coefficients as high as Pearson's $r = \pm 0.8$ are common in the literature (Elith *et al.*, 2006, 2010), and indeed, (multi)collinearity is considered as a serious challenge to most of the correlative modelling methods (Graham, 2003; Freckleton, 2011; Dormann *et al.*, 2013).

Recently, Zurell *et al.* (2012) proposed an environmental overlap mask that takes into account the novel combination of variables. However, this method provides no quantitative measure of projection uncertainty, but it simply classifies the test data (e.g. projection space) into analogous and novel categories and gives no information about the magnitude of dissimilarity (or similarity). Results obtained from this method can also be affected by the number of 'bins' selected by the user, which is a matter of subjectivity rather than statistical robustness (e.g. a bin number of one gives the same outputs as for MESS maps; see Zurell *et al.*, 2012, Supporting Information).

The reason that ignoring the correlation structure of the data can lead to erroneous conclusions about outliers can be appreciated from the following hypothetical example. Figure 1 shows the climate space of two hypothetical variables with a correlation coefficient of $r = 0.76$. The solid

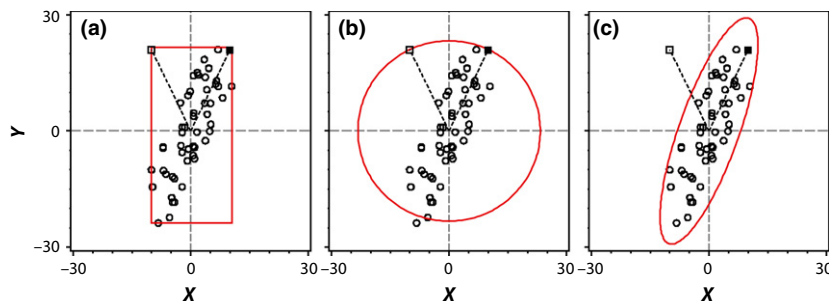


Figure 1 Environmental space of two hypothetical variables with a Pearson's correlation coefficient of $r = 0.76$. Solid square (Z1) and open square (Z2) symbols are two arbitrary points representing novel/outlier observations. Clearly, Z2 is further removed from the cloud of points than Z1 (see text). (a) Outermost data are used to delineate the rectilinear envelope; this is the method used by the MESS tool of MAXENT (Elith *et al.*, 2010) to identify novel points. The two putative novel points (Z1 and Z2) are therefore deemed identical as outliers as they both are located on the boundaries of envelope. (b) Under a spherical distribution (i.e. assuming the two variables are orthogonal), the two points (Z1 and Z2) are found again to be equally extreme because their distances from the centre of data are the same and equal to the circle radius. (c) Plotting an oblique ellipse which accounts for the correlation (and variance) between the variables enables discrimination between the two points in terms of extremity.

and open squares are two arbitrary points at coordinates (Z1: [10, 21]) and (Z2: [-10, 21]), respectively, which are suspected to be outliers (or novel observations in an SDM). To test the genuineness of a point, the MESS tool of MAXENT (Elith *et al.*, 2010) uses the outermost data to draw the boundaries of a climate envelope; if the point is outside this 'box' (or rectilinear hypervolume in the case of > 2 variables), it is considered a novel point (Fig. 1a). Using this scheme, the two points (Z1 and Z2) are identified as being equally extreme because they both touch the boundaries of the envelope, and indeed, the MESS calculations give a similarity value of 0 for both of them. Another approach is to use the Euclidean distance; again, it turns out that the two points are equally extreme as they have the same distance (radius) from the centre of data. Clearly, using a rectilinear (e.g. the MESS tool and Platts *et al.*, 2008; Fig. 1a) or spherical distribution (Fig. 1b) appears unreasonable ways to define the extent to which these observations are outliers from the rest of the data. That is, the two points are not equal; their directions of departure from the space of the other points are not the same. Point Z1 (solid square) lies in the same direction as the bivariate distribution of data and is thus less of an outlier. In contrast, the direction of point Z2 (open square) deviates from the pattern of scattered data and shows a different correlation structure. The distance of point Z2 could better be appreciated once the ellipse (with 95% confidence level) is estimated for the data (Fig. 1c).

This simple example (Fig. 1) illustrates the need for an improved method for detecting environmental novelty that accounts for the correlation between variables. As Dormann (2007) points out, 'far more problematic is the projection of a species' range under a combination of environmental conditions it has never encountered before'. Furthermore, while highlighting areas for further development after outlining their MESS map tool, Elith *et al.* (2010) state 'there remains

much scope for providing relevant tools for visualizing changes in correlations [between variables]'.

In this study, we propose a new method for the detection and quantification of novel environments that addresses this need by quantifying extrapolation due to covariate range and correlation change when projecting SDMs. Our method uses the Mahalanobis distance metric (Mahalanobis, 1936) to detect novel covariate combinations (correlations) that can occur within the univariate range of covariates. Furthermore, we provide additional tools for identifying and mapping the covariates that cause novelty. Finally, we demonstrate how the output can be visualized on a map, using an example model of the shrub *Acacia cyclops* (Fabaceae), a native to south-western Australia and an invasive alien elsewhere in the world. Our new extrapolation detection (ExDet) tool has been prepared as a stand-alone software package to integrate with most correlative SDM techniques and is freely available for download from <http://www.climond.org/exdet>.

METHODS

Defining novelty in model extrapolation

There is solid ecological support for identifying two types of novelty (i.e. extrapolation; e.g. Thuiller *et al.*, 2004b; Dormann, 2007; Elith *et al.*, 2010; Zurell *et al.*, 2012) when projecting SDMs: points that are novel because they are *outside the range of individual covariates* (defined here as Type 1 novelty; NT1; solid black circles in Fig. 2) and points that are *within the univariate range but constitute novel combinations between covariates* (defined here as Type 2 novelty; NT2; solid blue circles in Fig. 2). A third scenario is that points are both outside the univariate range and forming novel combinations; however, as this type of novelty is embedded in Type 1 novelty, we are not considering it here.

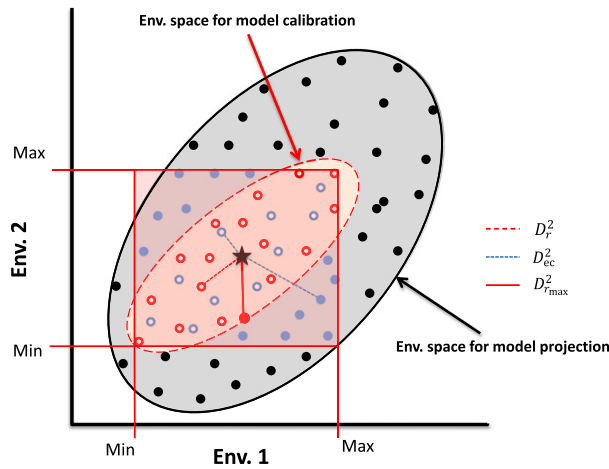


Figure 2 Schematic presentation of the ExDet tool for the detection and quantification of extrapolation in correlative SDMs using two hypothetical environmental variables. Red open circles represent the distribution records that define the sampled environmental space (small pink oval) used for model calibration, and the red rectangle shows the univariate coverage of this space. Black and blue solid circles and blue open circles represent grids on which the model is going to be projected and thus define the projection environmental space (large grey oval). All grids from the projection domain that are outside the rectangle (black solid circles: Type 1 novelty; NT1) are trimmed, and the degree of their dissimilarity is calculated using the NT1 component of the tool (equations 1 and 2). The remaining projection grids may represent covariate combinations not captured (blue solid circles: Type 2 novelty; NT2) or captured (blue open circles) in the sampled environments of calibration data. For each grid within the rectangle, the Mahalanobis distance (D^2) is calculated with respect to the centre of the environmental space of the calibration data (black solid asterisk) using equation 4 (for calibration data) or equation 6 (for projection data). The maximum distance found in the calibration data ($D_{r_{max}}^2$: red line with corresponding point shown as solid red circle) is then used to delineate the 'boundary' of data (dashed red ellipse). If the Mahalanobis distance of a point in the projection space (D_{ei}^2) is larger than the $D_{r_{max}}^2$, that point represents a novel environment (blue solid circle). For both analogous and novel points, a multivariate combination novelty index (NT2) is then calculated using equation 5. SDM, species distribution models.

The ExDet tool

Extrapolation detection

A rectilinear technique (red rectangle in Fig. 2) is first used to detect and quantify the dissimilarity of NT1 points (i.e. outside the univariate range). For any projection cells, a univariate distance (UD_{ij}) is calculated:

$$UD_{ij} = \frac{\min\{P_{ij} - \min(r_j), \max(r_j) - P_{ij}, 0\}}{\max(r_j) - \min(r_j)} \quad (1)$$

where P_{ij} is the value of grid cell i from the projection data (P) over the covariate j and $\min(r_j)$ and $\max(r_j)$ are the

minimum and maximum values for the same covariate, j , over the reference data (r), respectively. The above formula returns negative values for the projection points that are lower than the minimum or larger than the maximum ranges of the reference covariates (solid black circles in Fig. 2) and yields zero values for points within the range of reference data. The UD_{ij} values for a grid cell are then summed over all covariates ($j = 1 \dots p$, where p is the number of covariates) to give the degree of univariate novelty, NT1, for the respective point:

$$NT1_i = \sum_{j=1}^{j=p} UD_{ij} \quad (2)$$

NT1 ranges from infinite negative values to zero where zero indicates no extrapolation beyond the univariate coverage of reference data. The selection of reference data may depend on the modelling approach and software used. For example, if modelling with MAXENT (Phillips *et al.*, 2006), a viable option is to use the background as the reference set, while for envelope style models (e.g. BIOCLIM; Busby, 1986), one may use the actual presence records (e.g. from the native range of a species that has been introduced elsewhere).

The remaining points (blue open/solid circles in Fig. 2) that are inside the range (i.e. with $NT1 = 0$) are tested for novel covariate combinations (i.e. potential Type 2 novelty) using the Mahalanobis distance metric (D^2). The Mahalanobis distance is a well-known metric for the detection of multivariate outliers (Rousseeuw & Van Zomeren, 1990; Barnett & Lewis, 1994). The detection of 'extrapolation' in an SDM is analogous to the identification of multivariate 'outliers' in statistics: in both cases, the question is how to assess whether an observation in a multidimensional space belongs to a sample population and we need to attempt to measure its 'distance' from the sample points. In general, the Mahalanobis D^2 represents a scale-invariant (dimensionless) measure of the distance between centroids (mean vector) of two groups (Legendre & Legendre, 2012). However, the concept can be extended to calculate the distance of an observation in a p -variable data set to the centroid of a distribution or the distance of a multidimensional point to another one (all pairwise distances). For a p -dimensional multivariate sample \mathbf{x}_i ($i = 1, \dots, n$), the Mahalanobis D^2 is defined as follows:

$$D^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{Z}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad (3)$$

where $\bar{\mathbf{x}}$ is the vector of means: $\bar{\mathbf{x}}' = (\bar{x}_1, \dots, \bar{x}_p)$ and \mathbf{Z} the covariance matrix. This metric, therefore, reflects the correlation between variables by incorporating the covariance matrix into the calculation of distance values. The metric can have a minimum value of zero with unbounded upper limit.

Now, let $\bar{\mathbf{r}}$ be the mean vector of p variables (e.g. climate covariates) over a reference point set (red open circles in Fig. 2) with the corresponding variance–covariance matrix \mathbf{C} .

We calculate the distance of any point (D_r^2) in the reference set from the centre of data (black star in Fig. 2):

$$D_r^2 = (\mathbf{r}_i - \bar{\mathbf{r}})' \mathbf{C}^{-1} (\mathbf{r}_i - \bar{\mathbf{r}}) \quad (4)$$

The maximum distance found in the reference data ($D_{r_{\max}}^2$: red line with corresponding point shown as solid red circle in Fig. 2) is then used to delineate the ‘boundary’ or the ‘edge’ of data [dashed red ellipse; see Flury & Riedwyl (1988) for more details on calculating the ellipse]. It is important to note that all points on this dashed red ellipse have the same Mahalanobis distance as with $D_{r_{\max}}^2$. The novelty of points that we find to be within the univariate range, \mathbf{e}_i (blue open/solid circles in Fig. 2) is examined based on their position relative to this ellipse (or ellipsoid for $p > 2$). If a multivariate point, \mathbf{e}_i , is outside this ellipse, it is considered novel (NT2; solid blue circles in Fig. 2), while those inside the ellipse (open blue circles in Fig. 2) are reported as similar points. In general, a multivariate combination novelty index (NT2) is calculated based on its distance to the edge of data (ellipse) using the following formula:

$$\text{NT2}_i = \frac{D_{ei}^2}{D_{r_{\max}}^2} \quad (5)$$

where D_{ei}^2 is the Mahalanobis distance of \mathbf{e}_i from the centre of the reference data (i.e. $\bar{\mathbf{r}}$):

$$D_{ei}^2 = (\mathbf{e}_i - \bar{\mathbf{r}})' \mathbf{C}^{-1} (\mathbf{e}_i - \bar{\mathbf{r}}) \quad (6)$$

NT2 can range from zero up to unbounded positive values. NT2 values ranging from zero to one indicate similarity (in terms of both univariate range and multivariate combination), with values closer to zero being more similar. Values larger than one are indicative of novel combinations. Indeed, NT2 performs an additional test on points that have been reported as ‘similar’ when subjected to the NT1 test and returns NT2 values larger than one if these points show a different correlation structure.

It is worth clarifying that NT2 provides an indirect method of testing for changes in correlation because it is not possible to obtain a covariance matrix for a single point. Using D_e^2 , we initially assume that any new observation from projection data comes from a population with the same covariance structure as the known reference sample (here the background). However, if the D_e^2 of that point is larger than the maximum D_r^2 found in the reference data ($D_{r_{\max}}^2$), it would indicate that the questionable point does not belong to the same population because of a different covariance. A larger D_e^2 than expected (i.e. $D_e^2 > D_{r_{\max}}^2$), however, might also be the result of a difference in location too. As NT2 operates only on points that are within the range of reference data, we can conclude that the departure is mainly due to a different covariance (i.e. a different direction of the point relative to the axis of correlation in reference data) rather than a different location. Because NT2 is based on a linear correlation between covariates, the suitability of the method for nonlinear situations remains an issue open for future development.

To visualize the geographical extent of Type 1 novelty and Type 2 novelty, the mapped projection of an SDM can be accompanied or overlaid by the novelty data produced by the ExDet tool. This provides a clear indication of where the model is prone to extrapolation and the associated magnitude of novelty.

Assessing the contribution of covariates

The ExDet tool is also able to quantify the importance or contribution of each of the covariates to extrapolation in the form of a most influential covariate (MIC) metric. For a given point, a covariate that leads to the highest negative UD_{ij} (i.e. most extreme univariate range) is reported as the MIC leading to Type 1 novelty. For Type 1 novelty, our MIC metric gives the same results as MAXENT’s MoD (most dissimilar variable; Elith *et al.*, 2010). To identify the covariate driving Type 2 novelty, we make use of a unique characteristic of the Mahalanobis metric. That is, the Mahalanobis distance for any point will acquire its highest value if all covariates are included and will decrease as covariates are dropped sequentially. A covariate has a high contribution to the Type 2 novelty if the omission of that covariate (while keeping all other covariates in the model) makes a large reduction in the Mahalanobis distance (D_{ei}^2), as compared with the situation where all the covariates are included. The importance of a covariate p (IC_p) is therefore calculated based on the percentage reduction in Mahalanobis distance that it makes if dropped:

$$IC_p = 100 \times \frac{D_{\text{all}}^2 - D_{\text{all}-p}^2}{D_{\text{all}}^2} \quad (7)$$

where D_{all}^2 is the Mahalanobis distance for a point in the presence of all covariates and $D_{\text{all}-p}^2$ is the recalculated Mahalanobis distance when the covariate p has not been included. For any point, the MIC is the covariate which produces the highest IC_p value.

The exploration of dissimilar regions can be complemented by exploring the spatial distribution of the MICs within these regions. This MIC map helps to locate where any particular covariate has the most extreme univariate ranges (i.e. NT1) or its highest contribution to the largest correlation distortion (i.e. NT2).

The software package

Development of the extrapolation detection (ExDet) tool was performed using MATLAB (R2013b; the MathWorks Inc., Natick, MA, USA). In addition to the extrapolation detection and MIC map components, we include further data interrogation tools, such as data plotting for model covariates and projections. The tool comes as a platform versatile stand-alone software package available for free download from <http://www.climond.org/exdet>.

Example application

To demonstrate the value of the ExDet tool as an improved way to identify extrapolation, we apply it to a correlative model of *A. cyclops* A.Cunn. ex G.Don (Fabaceae). *Acacia cyclops* is a shrub native to the south-west of Western Australia and South Australia that has invasive alien populations in eastern parts of Australia and that has become a successful invader in other regions of the world (e.g. South Africa, the Iberian Peninsula and California; Fig. 3 in Webber *et al.*, 2011; Richardson & Rejmánek, 2011). We analysed the same distribution data used by Webber *et al.* (2011) for modelling the potential distribution of *A. cyclops*. Details on species' ecology as well as data collation, processing and quality control can be found in Webber *et al.* (2011). We built our model using native records of *A. cyclops*, given the problems that can be associated with combining native and alien data in discriminatory correlative models (Webber *et al.*, 2011), but used alien range data to inform our model interrogation. In particular, we focused on model performance and extrapolation behaviour in South Africa, where the *A. cyclops* invasion has been characterized by high propagule pressure

during introductions, has had considerable time to spread and is extremely well characterized (i.e. extensive collection of records with minimal sampling bias) across the entire known range (Richardson *et al.*, 2011; Webber *et al.*, 2011). That is, it is more likely to be closer to equilibrium and less likely to be affected by sampling bias and, therefore, represents an excellent opportunity to explore questions relating to model extrapolation.

Six bioclimatic covariates were chosen a priori that are reflective of average and extreme hydrothermal conditions: annual mean temperature (Bio1), maximum temperature of the warmest month (Bio5), minimum temperature of the coldest month (Bio6), annual precipitation (Bio12), precipitation of the wettest month (Bio13) and precipitation of the driest month (Bio14). We use the *A. cyclops* model here as an example only and recognize that the different choice of covariates between this model (based on generic climatic drivers of plant distribution) and that of Webber *et al.* (2011; based on ecological stress factors in the native range of *A. cyclops*) may lead to some differences in modelled projections. Global climate data (in the form of ASCII grids) of the chosen covariates were obtained from the CliMond data

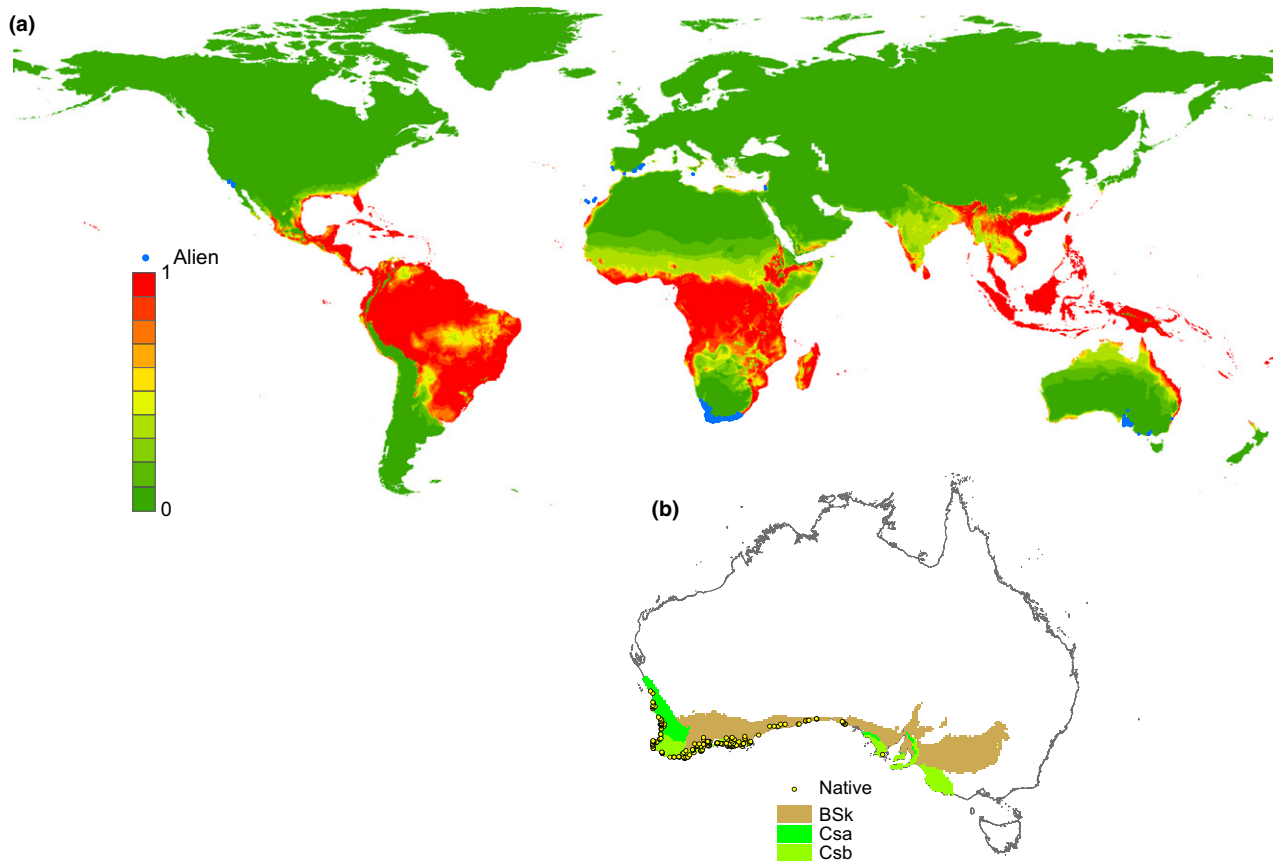


Figure 3 MAXENT's projection for relative habitat suitability (0–1) of *Acacia cyclops* (a) using native records of occurrence data (yellow circles in b) and Köppen–Geiger classification zones in the background (b). The background region was defined as all Köppen–Geiger classes (BSk: cold semi-arid; Csa and Csb: mediterranean) in Australia with at least one native record of *A. cyclops*, following Webber *et al.* (2011). Alien records of *A. cyclops* (blue circles in a) are found in eastern Australia and elsewhere in the world.

set (version 1.1; Kriticos *et al.*, 2012) at a resolution of 10' for current climate conditions (1975H).

We used MAXENT (version 3.3.3e; Phillips *et al.*, 2006) for modelling the global potential distribution of *A. cyclops*. We used default settings with the exception of restricting the model building to hinge features as simpler response curves are preferred for modelling invasive alien species (Elith *et al.*, 2010). We also selected 'create response curves' (response curves were clamped) and 'do jackknife' options. Following Webber *et al.* (2011), we used Köppen–Geiger classification zones (provided with the CliMond 10' 1975H data, Kriticos *et al.*, 2012) for defining the background required by MAXENT as a sample of the environments of a reference set against which to compare the presence records (Phillips *et al.*, 2006). The background region included polygons in Australia for all Köppen–Geiger classes containing at least one native presence record of *A. cyclops* (i.e. BSk, Csa, Csb; Fig. 3b). The same background sample was also used as the reference data set for calculating the Mahalanobis-based indices described above. The model was projected to the world.

Two measures of model performance were quantified: (1) the area under the curve statistic (AUC; Fielding & Bell, 1997) for native range records and (2) modelled sensitivity and specificity for alien range records, using the lowest presence threshold (LPT) for native range records (0.089; Webber *et al.*, 2011). Alien records that were located within the background region were excluded from these latter tests ($n = 127$). We used a one-tailed binomial test ($P(X \geq n)$) to test the probability (P) that the overall model sensitivity score could be achieved by chance alone, given the modelled prevalence. We also hypothesized that if there was putative ecological significance behind geographical components identified as similar, NT1 and NT2, that distribution records would be more commonly associated with similar regions and less commonly associated with NT1 and NT2 regions. This structure assumes that if we have equilibrium in the native range, this should be represented by occupancy of similar covariate space in the alien range. Binomial tests based on the proportion of the projection cells in each

region as P (probability of success) should therefore differ from random following $P(X > n)$ for similar regions and $P(X < n)$ for NT1 and NT2 regions. Finally, the uncertainty associated with the model projection, due to extrapolation to novel covariate space, was explored using the ExDet tool described above.

RESULTS

Maximum entropy modelling of *A. cyclops* attained a high value of the area under the curve statistic (AUC = 0.92; Fielding & Bell, 1997) when measured for the native range records, suggesting that the projection matched well the distribution records on which it was trained (Araújo *et al.*, 2005).

The model projection for the potential global distribution of *A. cyclops* showed that a large proportion of the global land mass at the high latitudes (North Hemisphere) is unsuitable for the species (Fig. 3a). Areas of high projected suitability for *A. cyclops* were in mediterranean-type climates and also tropical and subtropical regions. Modelled prevalence and sensitivity for the global projection using all distribution records were 0.249 and 0.909, respectively, the latter being highly significant ($P < 0.001$; Table 1). When this global projection was broken down into classes of similar covariate space, Type 1 novelty (NT1) and Type 2 novelty (NT2; relative to the background), the association of alien distribution records showed considerable variation between classes of covariate space (Table 1). Relative to what would be expected randomly, significantly more records were located in similar covariate space, and significantly, fewer records were located in NT1 regions ($P < 0.001$), whereas the number of records in NT2 regions was not different from a random distribution ($P = 0.293$; Table 1).

A preliminary analysis on the correlation structure of the data sets showed that the correlation coefficients among covariates varied considerably between the reference and global projection data (Table 2). For some pairs (eight of 15), even the sign of correlation coefficients changed across the two

Table 1 Goodness-of-fit measures and the alignment between distribution records and projection components for the MAXENT model of *Acacia cyclops*

Region	Component	Proportion of projection grids	Proportion of alien records	Modelled prevalence	Sensitivity	P-value
Global	Total			0.250	0.898	< 0.001
	Similar	0.007	0.641	0.208	0.897	< 0.001
	NT1	0.981	0.349	0.249	0.909	< 0.001
	NT2	0.012	0.010	0.357	0.571	0.293
South Africa*	Total			0.248	0.879	< 0.001
	Similar	0.112	0.705	0.420	0.861	< 0.001
	NT1	0.728	0.287	0.178	0.927	< 0.001
	NT2	0.161	0.008	0.446	0.750	< 0.001

*Within the mainland political boundaries (i.e. excluding Lesotho and Swaziland). Statistical tests (binomial), based on the proportion of projection cells in each region as P (probability of success), varied according to the component of the projection being assessed – 'total': $P(X \geq n)$ for sensitivity; similar: $P(X > n)$; NT1 and NT2: $P(X < n)$.

Table 2 The Pearson's correlation coefficients between six BIOCLIM covariates within the background (upper triangle, $n = 2862$) and projection data (lower triangle, $n = 563,019$) used for modelling the potential distribution of *Acacia cyclops*

	Bio1	Bio5	Bio6	Bio12	Bio13	Bio14
Bio1	—	0.65	0.17	−0.45	−0.25	−0.35
Bio5	0.90	—	−0.49	−0.49	−0.37	−0.01
Bio6	0.97	0.77	—	0.39	0.53	−0.57
Bio12	0.38	0.13	0.49	—	0.94	0.10
Bio13	0.46	0.24	0.53	0.90	—	−0.20
Bio14	0.06	−0.13	0.17	0.71	0.40	—

Annual mean temperature (Bio1), maximum temperature of the warmest month (Bio5), minimum temperature of the coldest month (Bio6), annual precipitation (Bio12), precipitation of the wettest month (Bio13), precipitation of the driest month (Bio14).

data sets. For example, the correlation between covariates Bio5 and Bio6 in the reference data of *A. cyclops* was negative ($r = -0.49$), but these covariates were positively correlated in the global projection data ($r = 0.77$).

As evident in the NT1 component of the ExDet map (Fig. 4), much of the global projection (Fig. 3a) was outside the univariate range of climate covariates (red in Fig. 4). The remaining projection areas that were within the univariate range (i.e. with no NT1) were further examined with the NT2 component of the ExDet tool. The NT2 component revealed

that more than half of these grids (6617 of 10,785) form novel combinations of climatic covariates that were not present in the model reference domain (blue in Fig. 4). For example, NT1 failed to detect the summer rainfall zones in South Africa as novel because these areas were within the range of climate covariates (Fig. 4 inset). However, for these regions, the correlation between temperature and precipitation can be positive as opposed to a mediterranean climate (i.e. similar to the reference data) where most of the precipitation falls in the winter. The NT2 component, however, was able to unravel this novel combination of covariates and yielded negative similarity values for these grids (Fig. 4 inset).

In South Africa, where the *A. cyclops* invasion is well characterized (see Methods), modelled prevalence and sensitivity were 0.248 and 0.879, respectively, the latter being highly significant ($P < 0.001$; Table 1; Fig. 4 inset). Regions with covariate space similar to the background represented just 11% of the country but contained 71% of the distribution records, which was significantly more than what could be expected by chance ($P < 0.001$; Table 1). In contrast, NT1 and NT2 regions contained significantly fewer records than what could be expected by chance ($P < 0.001$; Table 1). For example, the NT2 regions represent 16% of the area of South Africa but contained $< 1\%$ of the distribution records (Table 1; Fig. 4 inset).

Examination of the model response curves (see Fig. S1 in Supporting Information) suggested that there was no reduction in projected suitability at either end of the precipitation-based

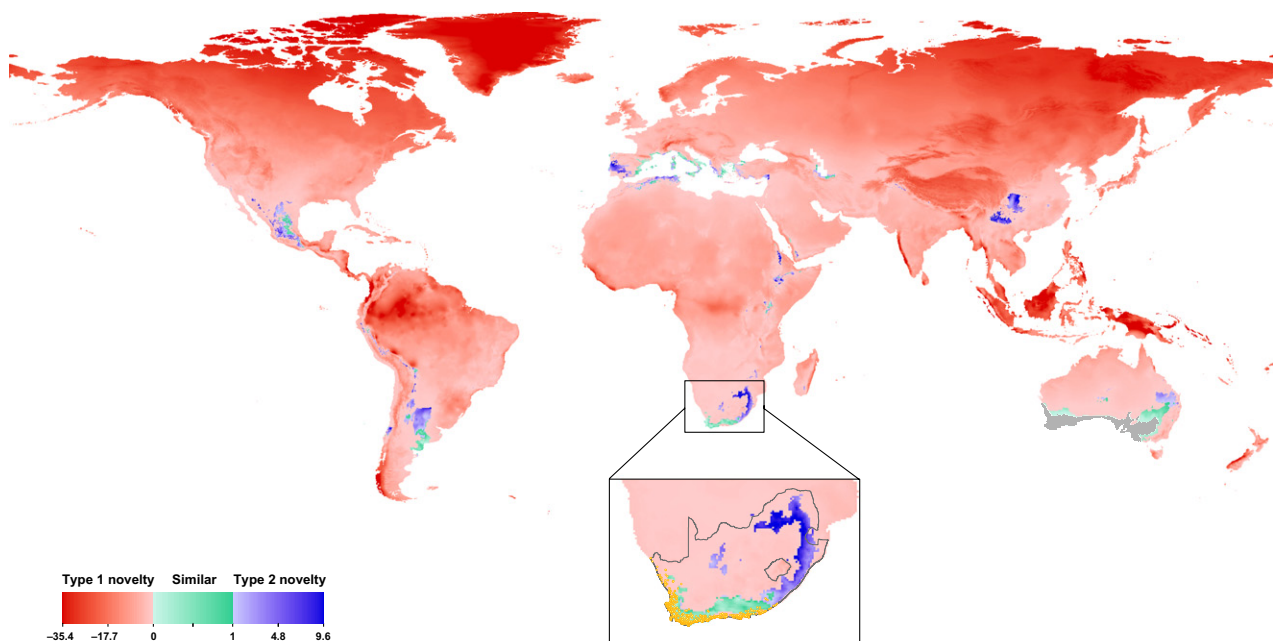
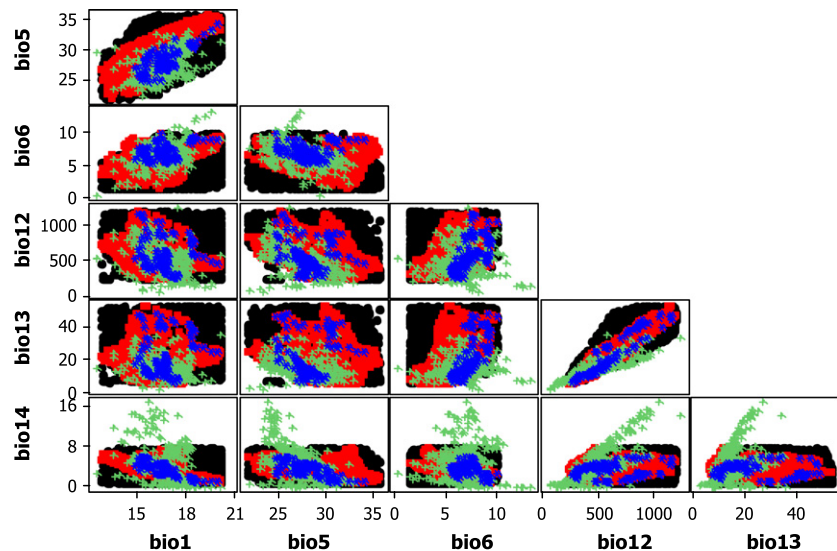


Figure 4 The ExDet map showing the quantification of extrapolated projection areas for *Acacia cyclops* quantified using the ExDet tool. The Type 1 novelty (NT1; equation 2; graded red) component indicates areas in the projection space with at least one climate covariate outside the univariate range of reference data. The Type 2 novelty (NT2; equation 5; graded blue) component indicates areas that are within the univariate coverage of reference data but represent non-analogous covariate combinations. Areas with NT2 ranging from 0 to 1 (graded green) are similar to the reference data as they both fall within the range of reference covariates and capture the same covariate combinations (equation 5). The area used as background is shown in grey in South Australia. The inset map (South Africa) highlights the low number of naturalized alien distribution records for *A. cyclops* in NT2 regions.

Figure 5 Scatter plots of six BIOCLIM variables for the *Acacia cyclops* model for the full global data set. Black circles show the subset of grids from the region of projection that were within the univariate range of calibration data but identified as novel by the NT2 component (equation 5). The novelty of these grids cannot be detected by the NT1 component (equation 2) because they are inside the univariate coverage of the model background data (red squares). Native (blue asterisk) and alien (green cross) global distribution records for the modelled species, *A. cyclops*, are overlaid. See Table 2 legend for BIOCLIM variable explanations.



covariates, while curve closure for temperature-based covariates happened only at one end of the curve and generally beyond the range experienced in most regions with high projected suitability. Taken together, this model interrogation suggests that the ecological plausibility of the projected suitability is likely to be highly questionable for *A. cyclops* (Fig. 3) in NT1 extrapolation space.

The extra insight provided by the NT2 component of our ExDet tool in revealing areas of Type 2 novelty became more apparent when geographical space (Fig. 4) was translated to environment space (Fig. 5) for the global projection. All black solid circles (6617 data points) that were not identified as novel based on the NT1 component, because of being within the range of reference data (red square), were found to be novel when examined by the NT2 component of ExDet tool (Fig. 5). Of note is the significant change in modelled covariate space that *A. cyclops* occupies in its alien range (green plus symbols in Fig. 5), relative to the native range. This covariate space is beyond the range and/or located within different combinations of covariates when compared with the native range of species.

Based on the MIC map for *A. cyclops*, the novel univariate range (i.e. NT1) for most areas was found to be driven by the minimum temperature of the coldest month (Bio6; Fig. 6). Bio6 was also the most frequent climatic variable resulting in novel combinations of covariates (i.e. NT2; Fig. 6). Annual mean temperature (Bio1) and annual precipitation (Bio12) were also major influential variables based on geographical extent (Fig. 6).

DISCUSSION

Almost all SDM projections to new regions and times may undergo some degree of extrapolation because not all the environmental ranges and possible combinations can be captured by the reference data sample. Indeed, a point can be novel in multivariate space even if it is within the range of any one univariate covariate (Fig. 4). Different methods have

been applied for comparisons of environmental space, each having their own advantages and disadvantages. The standardized Euclidean distance method can visualize dissimilar areas, but the method assumes sphericity (see Fig. 1b) even though climate covariates are rarely orthogonal. The MESS tool coded in MAXENT (Elith *et al.*, 2010) uses a rectilinear technique (Fig. 1a) for the detection of extrapolation, despite the fact that the climate envelope is often obliquely elliptic (Farber & Kadmon, 2003). Recently, Zurell *et al.* (2012) explored the problem of novel multivariate combinations by extending the idea of the MESS map into an environmental overlap mask (*sensu* Zurell *et al.*, 2012). Although promising, the resultant output is binary, and it does not measure the size of 'novelty' and the method is subjective.

By identifying both Type 1 novelty and Type 2 novelty, our ExDet tool can effectively identify novel multivariate combinations and quantify the magnitude of this novelty by implementing the Mahalanobis metric as a measure of environmental novelty/similarity. The method is functional for both orthogonal (either spherical or rectangle shape) and correlated (elliptical shape) covariates: mathematically, the Mahalanobis distance reduces to a standardized Euclidean distance when the covariance between variables approaches zero (i.e. variables are orthogonal to each other). The Mahalanobis metric has been successfully used as a presence-only method for habitat suitability modelling (e.g. Clark *et al.*, 1993; Farber & Kadmon, 2003; Thatcher *et al.*, 2006), and here, we add to this by applying it as an extrapolation detection tool.

There are several advantages in using the Mahalanobis metric for exploring extrapolation by SDMs. First, the metric accounts for the fact that the covariates in SDMs are correlated and may differ in variances. Second, the differences in dispersions between variables are accounted for as variables are standardized with respect to their standard deviations. Third, the metric is robust to departure from multivariate normality (Clark *et al.*, 1993); if normality is met (or

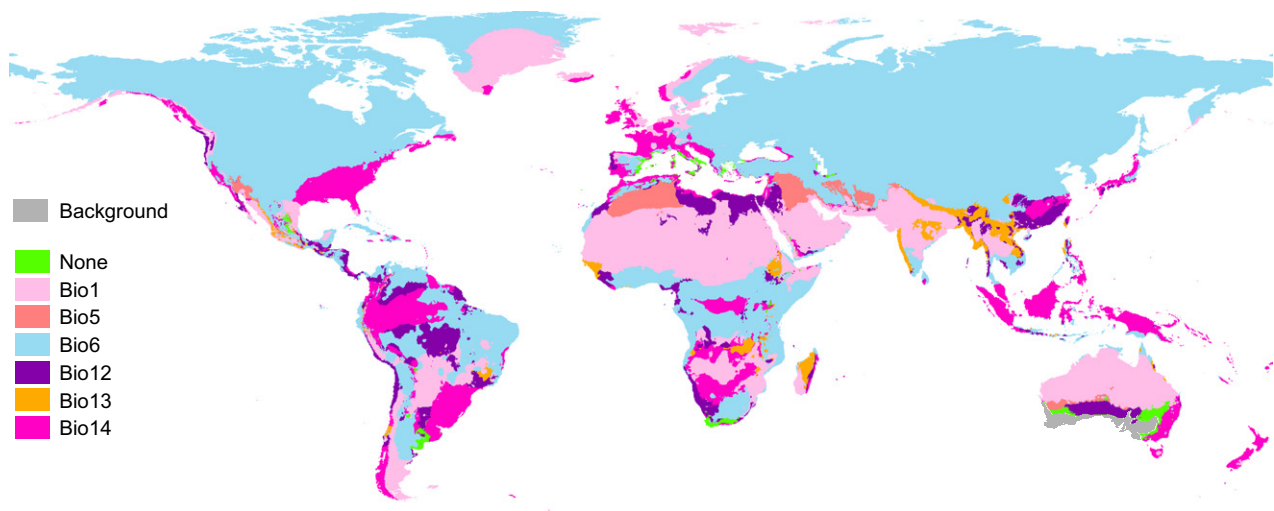


Figure 6 Most influential covariate (MIC, equation 7) map for the *Acacia cyclops* model. The spatial distribution of the most dissimilar covariate is shown in terms of their contribution to both Type 1 novelty and Type 2 novelty. Green areas have no covariate outside the coverage of reference data or with non-analogous combinations. The area used as background is shown in grey in South Australia. See Table 2 legend for BIOCLIM variable explanations.

achieved by some transformation), the bonus is that the chi-square distribution can be used to test whether a point is statistically novel. Fourth, the resultant MIC map can further help to determine which variable is most influential in driving dissimilarity for any region of the projection. This aspect is not considered in the method of Zurell *et al.* (2012), and the equivalent approach of Elith *et al.* (2010) is only based on the range of variables, not their correlation structure. MIC may also help prioritize the candidate variables more strategically (e.g. by choosing more transferable/less dissimilar covariates). When selecting covariates, however, other factors such as ecological relevance and the quality of fit could be more important than transferability. Finally, the use of Mahalanobis metric, which implies (but is not restricted to) an elliptical shape, is in agreement with niche theory (Whittaker & Levin, 1975) and Shelford's Law of Tolerance (Shelford, 1931). That is, species will be unevenly distributed along environmental gradients (e.g. model covariates), with a tendency towards a covariate optimum, along with minima and maxima beyond which they cannot exist. Similar theory has been used to underpin the use of the Mahalanobis metric as a way to undertake species distribution modelling (Farber & Kadmon, 2003).

Although the ExDet tool developed here allows for better awareness of where models are extrapolating, it is important to reinforce that model interpretation in these novel regions is fraught with danger (Elith *et al.*, 2010). Of great concern for correlative SDMs is that modellers and end-users of the projections are particularly interested in the ecological implications of the novel component of model projections (van Klinken *et al.*, 2009; Venette *et al.*, 2010; Webber *et al.*, 2012). For example, as shown here for *A. cyclops*, and elsewhere for other species (e.g. Broennimann *et al.*, 2007; Gallagher *et al.*, 2010; Thompson *et al.*, 2011), a shift in modelled

covariate space, relative to the native range, is common for introduced species. Such a difference could represent a true niche shift, provided that the underlying driver of range boundaries is captured by the modelled covariates. Alternatively, what appears to be a change in niche space may be a consequence of a change in correlation structure of the modelled covariates. Regardless of the underlying mechanism, novel climate space can be due to both the exceeding of covariate range (Type 1 novelty) and a change in the correlation between covariates (Type 2 novelty; Figs 4 & 5).

Up until now, Type 2 novelty has not been a focus for model interrogation when considering extrapolation issues. Even so, such interaction between covariates can have significant implications for species range limits. Frequently encountered situations include winter- versus summer-dominant rainfall or humidity-associated variation. In the *A. cyclops* example modelled here, the South African distribution revealed significantly fewer records for the shrub in NT2 regions, relative to what could be expected by chance (Fig. 4 inset; Table 1). The *A. cyclops* invasion in South Africa is unusual in that it is more likely to be closer to equilibrium and less likely to be affected by sampling bias. These results therefore provide putative evidence that the detected patterns may have ecological underpinnings that influence the potential range of *A. cyclops*. We recognize that an increased awareness of the properties of novel bioclimatic space does not always correspond with a better understanding of the range limiting factors for species distributions. Although there are methods for improving the performance of correlative SDMs in novel bioclimatic space (e.g. Elith & Leathwick, 2009), ecologically plausible extrapolations with these techniques remain an elusive goal.

Model extrapolation seems to be an inescapable part of many questions that correlative SDMs are applied to in the current literature. Given that correlative SDMs are likely to

remain an extremely popular tool to help identify the current and potential future distribution of species, extrapolation detecting tools are an essential component of these modelling systems to avoid erroneous inferences. The ExDet tool we introduce provides further insight that is lacking from tools currently available as it addresses the problem of incorporating and mapping novelty relating to both covariate range (NT1) and correlation change (NT2). This deficiency has been appreciated (Dormann, 2007; Elith *et al.*, 2010) but until now has not been addressed in a manner functional for end-users of SDMs. Our method currently supports linearly correlated, quantitative variables, which are the most common data type used in SDMs. Further work would be needed to extend the applicability of the method to nonlinearly correlated or categorical variables. We caution, however, that even if we are better able to identify regions of model extrapolation, we must remain vigilant and approach model interrogation in these regions with the upmost care if any conclusion other than 'here be dragons' is to be drawn.

ACKNOWLEDGEMENTS

M.B.M. was supported by the University of Melbourne John McKenzie Post-Doctoral Research Fellowship. B.L.W. acknowledges support from the CSIRO Climate Adaptation and Biosecurity Flagships. We thank Brendan Wintle, Bob O'Hara, Carsten Dormann and two anonymous referees for valuable comments that improved the manuscript. We thank Ehsan Mahmoudi and Arash Zarghamnejad for programming assistance. We thank the co-authors of Webber *et al.* (2011) for sharing distribution data for *Acacia cyclops*, which was quality controlled as part of research funded by the Australian Department of Climate Change and Energy Efficiency (through the Bilateral Climate Change Partnership Program).

REFERENCES

- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**, 1527–1539.
- Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Barnett, V. & Lewis, T. (1994) *Outliers in statistical data*, 3rd edn. Wiley, Chichester.
- Broennimann, O., Treier, U.A., Müller-Schärer, H., Thuiller, W., Peterson, A. & Guisan, A. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters*, **10**, 701–709.
- Busby, J.R. (1986) A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. *Australian Journal of Ecology*, **11**, 1–7.
- Clark, J.D., Dunn, J.E. & Smith, K.G. (1993) A multivariate model of female black bear habitat use for a geographic information system. *The Journal of Wildlife Management*, **57**, 519–526.
- Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B. & Wood, S. (1998) Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**, 783–786.
- Dormann, C.F. (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, **8**, 387–397.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
- Ebeling, S.K., Welk, E., Auge, H. & Bruehlheide, H. (2008) Predicting the spread of an invasive plant: combining experiments and ecological niche model. *Ecography*, **31**, 709–719.
- 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**, 677–697.
- 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.
- Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.
- Farber, O. & Kadmon, R. (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, **160**, 115–130.
- 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.
- Fitzpatrick, M.C. & Hargrove, W.W. (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255–2261.
- Flury, B. & Riedwyl, H. (1988) *Multivariate statistics: a practical approach*. Chapman & Hall, Ltd, New York.
- Franklin, J. (2010) Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, **16**, 321–330.
- Freckleton, R.P. (2011) Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology*, **65**, 91–101.
- Gallagher, R.V., Beaumont, L.J., Hughes, L. & Leishman, M.R. (2010) Evidence for climatic niche and biome shifts between native and novel ranges in plant species introduced to Australia. *Journal of Ecology*, **98**, 790–799.
- Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, **84**, 2809–2815.

- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Jeschke, J.M. & Strayer, D.L. (2008) Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, **1134**, 1–24.
- van Klinken, R.D., Lawson, B.E. & Zalucki, M.P. (2009) Predicting invasions in Australia by a Neotropical shrub under climate change: the challenge of novel climates and parameter estimation. *Global Ecology and Biogeography*, **18**, 688–700.
- Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J. & Scott, J.K. (2012) CliMond: global high resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, **3**, 53–64.
- Kriticos, D.J., Le Maitre, D.C. & Webber, B.L. (2013) Essential elements of discourse for advancing the modelling of species' current and potential distributions. *Journal of Biogeography*, **40**, 608–611.
- Legendre, P. & Legendre, L. (2012) *Numerical ecology*, 3rd edn. Elsevier, Oxford.
- Mahalanobis, P.C. (1936) *On the generalized distance in statistics*. pp. 49–55. Proceedings of the National Institute of Sciences of India, New Delhi.
- Meentemeyer, R.K., Anacker, B.L., Mark, W. & Rizzo, D.M. (2008) Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications*, **18**, 377–390.
- Peterson, A.T. & Soberón, J. (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. *Natureza & Conservação*, **10**, 102–107.
- Peterson, T.A., Papeş, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, **30**, 550–560.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Platts, P.J., McClean, C.J., Lovett, J.C. & Marchant, R. (2008) Predicting tree distributions in an East African biodiversity hotspot: model selection, data bias and envelope uncertainty. *Ecological Modelling*, **218**, 121–134.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- Rhodes, J.R., Wiegand, T., McAlpine, C.A., Callaghan, J., Lunney, D., Bowen, M. & Possingham, H.P. (2006) Modeling species' distributions to improve conservation in semi-urban landscapes: koala case study. *Conservation Biology*, **20**, 449–459.
- Richardson, D.M. & Rejmánek, M. (2011) Trees and shrubs as invasive alien species – a global review. *Diversity and Distributions*, **17**, 788–809.
- Richardson, D.M., Carruthers, J., Hui, C., Impson, F.A.C., Miller, J.T., Robertson, M.P., Rouget, M., Le Roux, J.J. & Wilson, J.R.U. (2011) Human-mediated introductions of Australian acacias – a global experiment in biogeography. *Diversity and Distributions*, **17**, 771–787.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, **35**, 211–226.
- Rousseeuw, P.J. & Van Zomeren, B.C. (1990) Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, **85**, 633–639.
- Schorr, G., Holstein, N., Pearman, P., Guisan, A. & Kadereit, J. (2012) Integrating species distribution models (SDMs) and phylogeography for two species of Alpine Primula. *Ecology and Evolution*, **2**, 1260–1277.
- Scoble, J. & Lowe, A.J. (2010) A case for incorporating phylogeography and landscape genetics into species distribution modelling approaches to improve climate adaptation and conservation planning. *Diversity and Distributions*, **16**, 343–353.
- Shelford, V.E. (1931) Some concepts of bioecology. *Ecology*, **12**, 455–467.
- Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., Alexander, J.D., Wiens, J.A. & Root, T.L. (2009) Re-shuffling of species with climate disruption: a no-analog future for California birds? *PLoS One*, **4**, e6825.
- Strubbe, D. & Matthysen, E. (2009) Predicting the potential distribution of invasive ring-necked parakeets *Psittacula krameri* in northern Belgium using an ecological niche modelling approach. *Biological Invasions*, **11**, 497–513.
- Sutherst, R.W. & Bourne, A.S. (2009) Modelling non-equilibrium distributions of invasive species: a tale of two modelling paradigms. *Biological Invasions*, **11**, 1231–1237.
- Thatcher, C.A., Manen, F.T.V. & Clark, J.D. (2006) Identifying suitable sites for Florida panther reintroduction. *Journal of Wildlife Management*, **70**, 752–763.
- Thompson, G.D., Robertson, M.P., Webber, B.L., Richardson, D.M., Le Roux, J.J. & Wilson, J.R.U. (2011) Predicting the subspecific identity of invasive species using distribution models: *Acacia saligna* as an example. *Diversity and Distributions*, **17**, 1001–1014.
- Thuiller, W., Araújo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L. & Lavorel, S. (2004a) Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature*, **430**, 6995.
- Thuiller, W., Brotons, L., Araújo, M.B. & Lavorel, S. (2004b) Effects of restricting environmental range of data to project current and future species distributions. *Ecography*, **27**, 165–172.
- Václavík, T. & Meentemeyer, R.K. (2009) Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling*, **220**, 3248–3258.
- Venette, R.C., Kriticos, D.J., Magarey, R.D., Koch, F.H., Baker, R.H.A., Worner, S.P., Gómez Raboteaux, N.N., McKenney, D.W., Dobesberger, E.J., Yemshanov, D., de Barro, P.J., Hutchison, W.D., Fowler, G., Kalaris, T.M. &

- Pedlar, J. (2010) Pest risk maps for invasive alien species: a roadmap for improvement. *BioScience*, **60**, 349–362.
- Webber, B.L., Yates, C.J., Le Maitre, D.C., Scott, J.K., Kriticos, D.J., Ota, N., McNeill, A., Le Roux, J.J. & Midgley, G.F. (2011) Modelling horses for novel climate courses: insights from projecting potential distributions of native and alien Australian acacias with correlative and mechanistic models. *Diversity and Distributions*, **17**, 978–1000.
- Webber, B.L., Le Maitre, D.C. & Kriticos, D.J. (2012) Comment on “Climatic niche shifts are rare among terrestrial plant invaders”. *Science*, **338**, 193.
- Whittaker, R.H. & Levin, S.A. (1975) *Theory and application. Benchmark papers in ecology*. Dowden, Hutchinson and Ross, Stroudsburg.
- Williams, J.W. & Jackson, S.T. (2007) Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475–482.
- Williams, J.W., Jackson, S.T. & Kutzbach, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences USA*, **104**, 5738–5742.
- Zurell, D., Elith, J. & Schröder, B. (2012) Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions*, **18**, 628–634.

SUPPORTING INFORMATION

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

Figure S1 Bioclimatic covariate response curves for *Acacia cyclops*.

BIOSKETCHES

Mohsen Mesgaran is a plant ecophysiologicalist who enjoys linking ecological patterns and physiological processes by implementing mathematical models. His current research focuses on coastal plant invaders and predicting their potential range expansion through the development of mechanistic models.

Roger Cousens is a plant population ecologist interested in processes that determine abundance and distribution of plant species and genotypes. He develops methodology and uses experimentation, survey and modelling to understand and predict how abundance and distribution change under different conditions and hence what action needs to be taken to manage populations.

Bruce Webber is a plant ecophysiologicalist exploring the effect of rapid global change on plant resource allocation and plant–ecosystem interactions. With interests in invasion science, conservation biology and food security, he uses a variety of methods to understand range shifts, landscape connectivity and plant fitness. He is passionate about improving best practice for bioclimatic modelling, particularly in regard to model extrapolation.

Author contributions: All authors conceived the ideas; M.B.M. led the tool development, B.L.W. collected the data, M.B.M. and B.L.W. analysed the data, and M.B.M. and B.L.W. led the writing.

Editor: Janet Franklin