



Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: [www.elsevier.com/locate/ecoinf](http://www.elsevier.com/locate/ecoinf)

## Predicting species distributions in new areas or time periods with alpha-shapes

César Capinha<sup>a,b,\*</sup>, Beatriz Pateiro-López<sup>c</sup>

<sup>a</sup> Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal

<sup>b</sup> Zoologisches Forschungsmuseum Alexander Koenig, Museumsmeile Bonn, Adenauerallee 160, 53113 Bonn, Germany

<sup>c</sup> Departamento de Estatística e Investigación Operativa, Faculdade de Matemáticas, Universidade de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, Spain

### ARTICLE INFO

#### Article history:

Received 8 February 2014

Received in revised form 30 May 2014

Accepted 1 June 2014

Available online xxx

#### Keywords:

Alpha-shapes  
Climate change  
Invasive species  
Niche modeling  
Presence-only  
*Xenopus laevis*

### ABSTRACT

Statistical models relating species distributions to environmental data are now commonly applied to predict where invasive species may become established or how range limits may shift under climate change. As species absences can originate from factors other than an unsuitable environment (e.g. dispersal constraints), the models that discriminate between occupied and unoccupied environments are likely to underestimate potential ranges. However, the techniques that “envelope” the occupied environments (i.e. profile techniques) usually rely on simple convex estimators (e.g. elliptical or rectangular shapes), which tend to overestimate these ranges. Here we describe alpha-shapes, a profile-type technique that relaxes the assumption of convexity. By using native range data for the invasive African clawed frog, we demonstrate how this technique can be used to model climatic envelopes of variable complexity. In particular, we compared predictions from an envelope maximizing discrimination between presences and absences, an envelope tightly enclosing all occupied climatic combinations (i.e. the minimum bounding envelope) and an “expert-based” generalization of the previous. In addition, we also use this technique to identify climatic combinations that are outside the climatic space of the study area (i.e. non-analog climates). The envelope accounting for the absences of the African clawed frog achieved a high discrimination ability (true skill statistics = 0.71), but failed to predict many of the areas in which the species occurs. Predictions based on the minimum bounding envelope encompassed all species occurrences while still providing a sharp delineation of its distribution range. The generalized version of the previous envelope also captured all occurrences, but predicted a wider extent of suitable areas. We also found that most parts of the world present climatic conditions that are non-analog to those of our study area. Although conceptually more suitable for predicting species distributions across space and time than presence-absence models, profile techniques are frequently overlooked because of their inability to fit flexible envelopes. Here, we demonstrate that alpha-shapes are a transparent and intuitive profile-type technique that has this flexibility.

© 2014 Elsevier B.V. All rights reserved.

### 1. Introduction

Correlative species distribution models (SDMs) (Elith and Leathwick, 2009; Guisan and Thuiller, 2005) are the most practical and popular approach for estimating the geographical distribution of species. Mainly used for interpolation in their early years (e.g. Carroll et al., 1999), these models are now also widely applied in a range of ecological and evolutionary research questions that imply extrapolating species distributions to new areas or time periods, such as anticipating biological invasions (Jiménez-Valverde et al., 2011), predicting species' past distributions (Nogués-Bravo, 2009) or range shifts in response to climate change (Thomas et al., 2004). In parallel, the statistics, based on which SDMs have been developed, have also diversified considerably: complex machine learning algorithms, regression-based models, classification

techniques, distance metrics, and environmental envelopes are all examples of techniques now commonly employed (e.g. Elith et al., 2006). Particularly interesting is the resilience of profile techniques, i.e. those exclusively using species occurrences. These correspond to some of the simplest and earliest statistics used for SDM, such as BIOCLIM (Busby, 1991), DOMAIN (Carpenter et al., 1993), or Mahalanobis distances (Farber and Kadmon, 2003), and despite their apparently lower predictive ability (Elith et al., 2006), they are still commonly used nowadays (e.g. Diniz-Filho et al., 2010; Munguía et al., 2012). In the following paragraphs, we briefly discuss some of the attractive properties of profile techniques, which make us believe that they will continue to be valuable in the future, particularly for predicting potential distributions in new spatial or temporal domains. Furthermore, we introduce the use of alpha-shapes (Edelsbrunner and Mücke, 1994; Edelsbrunner et al., 1983) for SDM, which is a profile-type technique that allows overcoming some of the most common limitations of this group of models.

Perhaps, the main attractiveness of profile techniques concerns their independence from absence data. While data on species presence are now widely available from museums or online databases (e.g. GBIF;

\* Corresponding author at: Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal. Tel.: +351 217500577.

E-mail address: [cesarcapinha@outlook.com](mailto:cesarcapinha@outlook.com) (C. Capinha).

<http://data.gbif.org>), absence data are, however, often inexistent. This poses a serious difficulty to the application of presence–absence models. A common workaround is to contrast species presence data with the environmental background, and although good results are apparently achievable (e.g. Barbet-Massin et al., 2012), the way in which the background is delimited can severely influence the predictions (Barbet-Massin et al., 2012; Capinha et al., 2011; Elith et al., 2010). Moreover, when predictions from presence–absence models are applied to new areas or time periods, it is imperative to consider the factors that are responsible for the absence of species. As the usual aim of SDM in this context is to find the suite of abiotic conditions suitable for species (e.g. Elith et al., 2010; Nogués-Bravo, 2009; Thomas et al., 2004), the use of absence records caused by non-abiotic factors such as dispersal barriers or competitors may lead to an underestimation of potential distributions (Jiménez-Valverde et al., 2008, 2011).

Another main feature in favor of profile techniques for extrapolation concerns their overall technical simplicity and consequent ease of interpretation. Model extrapolation is an undoubtedly risky task and, as such, should be preceded by a careful scrutiny of the species–environment relationships that are fitted by the models (e.g. Elith and Graham, 2009; Zurell et al., 2012). This implies that these relationships should be possible to visualize and not be overly complex (in order to be interpretable). This is, however, not always the case for presence–absence techniques, which often produce “opaque” non-parametric response functions composed of a large number of terms (e.g. Prasad et al., 2006). Moreover, as the researchers' own understanding about the species ecology may be crucial under extrapolation, it may also be desirable to opt for models allowing an increased “control” over parameterization. An important distinction between profile techniques and most of the presence–absence models is that the latter automatically assign weights to each predictor based on its power to differentiate occurrences from absences (or environmental background). This characteristic can, however, be undesirable under novel conditions because the relative importance of the predictors may not be conserved (Rödder and Lötters, 2010). Alternatively, profile techniques assume equal weights for all predictors, allowing researchers to select relevant predictors based on their own expertise.

Despite the potential advantages mentioned earlier, profile techniques also share several characteristics that can hinder their predictive ability, regardless of the spatial or temporal domain being investigated. Importantly, most of these techniques assume the shape of the species' realized niche (i.e. the set of suitable environmental conditions which is occupied) *a priori*. For instance, BIOCLIM fits a rectilinear envelope, while Mahalanobis distances fit elliptical shapes. The use of simplified geometries is an important drawback of these models, because it can lead to the inclusion (i.e. predict as suitable) of many environmental combinations that are not, in fact, occupied by the species (Elith and Burgman, 2003; Farber and Kadmon, 2003; see below). In addition, these methods are also sensitive to outliers and, in the case of distance metrics, to bias in the calculation of the species “optimum” from which distances are measured (Calenge et al., 2008; Rotenberry et al., 2006). Here, we discuss the application of alpha-shapes to predict species distributions across time or space. Alpha-shapes differ fundamentally from most of the previous profile techniques as they make no assumption about the shape of the niche. Instead, alpha-shapes attempt to capture the overall “shape” of the niche using a flexible envelope fitting procedure. Moreover, this envelope can be visualized, scrutinized, and readjusted, if necessary. In the following section, we describe what alpha-shapes are and exemplify how they can be applied for SDM using data from the African clawed frog (*Xenopus laevis*), a high-profile invasive species. We provide modeling guidance and show how to detect and measure the potential impact of outliers on the geographical projections. Finally, we also show that alpha-shapes can also be applied to delimit non-analog environments, a well-known source of uncertainty in this field of research (Fitzpatrick and Hargrove, 2009).

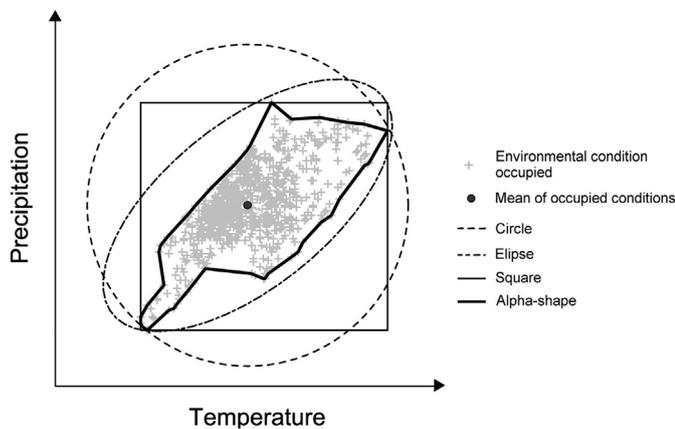
## 2. An introduction to alpha-shapes

The study of shape is a primary theme in many disciplines. To that end, different methodologies from fields such as geometry, statistics, and computer science have been proposed (for a review, see Zhang and Lu, 2004). Set estimation theory is a branch of statistics that deals with the problem of approximating an unknown set based on a random sample of points taken from it. The endeavor of SDMs can be addressed as a problem of set estimation theory, where the  $n$ -dimensional envelope that defines a species' niche (Hutchinson, 1957) is approximated based on the position of presence or presence–absence records in the environmental space. Usually, the problem of set estimation lies on the identification of a suitable estimator for the target set (see Cuevas and Fraiman, 2009). When the aim is to capture the shape of a set from a random sample of points (as in SDM), there are different geometrical structures that can be used as estimators. In many cases, the attempt relies on the use of simplified convex estimators such as the convex-hull or circular forms. However, often, the set is not strictly convex, and a convex estimator does not produce an effective reconstruction of its shape.

The alpha-convexity is a shape restriction that relaxes the assumption of convexity (Walther, 1997, 1999). An alpha-convex set is not necessarily connected and it can enclose cavities. To capture the shape of an alpha-convex set from a random sample of points, the alpha-convex hull of the sample is a suitable estimator (Pateiro-López and Rodríguez-Casal, 2013). The alpha-convex hull of a point set was first defined in the plane (Edelsbrunner et al., 1983) and then generalized to the 3D space (Edelsbrunner and Mücke, 1994). Intuitively, by using SDM terms, one can visualize the alpha-convex hull of a sample in 3D as what is left if we carve out all the environmental space encompassed by a spherical scoop of radius alpha ( $\alpha$ ) that does not include any of the species occurrences. The same concept applies in 2D, but by using a circular eraser. As  $\alpha$  tends to infinity, the alpha-convex hull tends to the convex hull of the sample. As  $\alpha$  decreases, the alpha-convex hull shrinks and cavities appear among the species occurrences. If we straighten the boundary of the alpha-convex hull – consisting of arcs of circles in 2D or spherical caps in 3D – we get the so-called alpha-shape. Alpha-shapes are subgraphs of the Delaunay triangulation of a point set where the parameter  $\alpha$  controls the simplices – i.e. tetrahedral (in 3D), triangles, edges, and points – that will or will not make part of an alpha-shape. The family of alpha-shapes formed along all possible values of  $\alpha$  is called the alpha-spectrum. Since the implementation of the theoretical algorithm to compute alpha-shapes, many applications of this geometrical structure have been presented in the literature (e.g. Edelsbrunner, 2010).

### 2.1. Comparison of alpha-shapes with other profile-type techniques

Prediction of SDMs based on profile-type techniques typically relies on a multidimensional geometric construct that bounds the environmental conditions occupied by the species. The environmental combinations falling within this envelope are classified as suitable, while those falling outside may be unsuitable. The most fundamental difference among the various techniques concerns the type of estimator used. For instance, in a 2D space, Euclidean distances assume that the species' niche will be circular, whereas Mahalanobis distances assume that it will be elliptical and BIOCLIM assumes that it will be rectilinear. The  $n$ -dimensional analogs of these shapes are also used in  $n$ -dimensional spaces. In Fig. 1, we compare, for a hypothetical species, the 2D environmental envelope generated by an alpha-shape and by the three techniques mentioned earlier. As demonstrated, all the methods that define the geometric shape of the niche *a priori* can only provide a crude representation of the point cloud and consequently enclose wide portions of environmental space that is not, in fact, used by the species. This tendency for overestimation is particularly well known for Euclidean distances and BIOCLIM (Elith and



**Fig. 1.** Comparison of the 2D climatic envelopes fitted by circular, elliptical, and rectilinear shapes and by alpha-shapes for a hypothetical species. The alpha-shape that is represented is just one of the many that could possibly be obtained for this point set—i.e. tighter or looser envelopes can be selected.

Burgman, 2003; Farber and Kadmon, 2003), but it can also be expected for Mahalanobis distances (Farber and Kadmon, 2003).

However, the ability of alpha-shapes to draw non-convex boundaries allowed the more precise delimitation of the point cloud (Fig. 1). Importantly, distinct alpha-shape envelopes could be obtained by varying the parameter  $\alpha$ . In other words, a tighter or more general fit could be selected if desired (see section “Model implementation and selection of  $\alpha$ ”).

Another potentially advantageous characteristic of alpha-shapes for SDM is that this method does not rely on the definition of an “optimal” set of conditions for the species—a requirement of most distance-based envelopes. Although intuitive, in practice, the use of such reference set can be problematic because its calculation—usually, the mean of occupied conditions—is sensitive to bias in species sampling effort and uneven availability of suitable conditions (Calenge et al., 2008). In addition, if a species was not able to disperse to the full range of suitable environmental combinations, it may not be possible to know its “true optimum” (Rotenberry et al., 2006). Thus, distances may be measured to a non-optimal point of the species niche, which will ultimately affect the reliability of equidistant contours.

Because of a high-dimensional complexity, to date, alpha-shapes are only implemented up to three dimensions. This can be a limitation for fine-scale SDMs, for which it is sometimes necessary to consider a large number of factors driving the species’ distributions (e.g. topography, the distribution of other species, and climate). However, this is less of a problem for models intended for new temporal or geographical domains. The general aim of these models is to capture species–environment relationships that are sufficiently general to be informative under new conditions and, for this to be achieved, overfitting should be carefully avoided. Overfitting can be caused by the 1) inclusion of irrelevant predictors causing the fit of labile relationships (Peterson, 2011) and 2) fitting of “noise,” i.e. spurious patterns that are interpreted as “legitimate” (Vaughan and Ormerod, 2005). For practical reasons, the solution for both cases converges on the use of a low-dimensional environmental space. Concerning the first cause, a low number of important predictors is always preferable to the use of a high-dimensional set of dubious relevance (Peterson, 2011). Regarding the second cause, a visual scrutiny in two or three dimensions remains one of the most practical approaches for detecting the generality of model fits (e.g. Elith and Graham, 2009; Zurell et al., 2012). If after accounting for overfitting, the set of predictors exceeds three dimensions, data reduction techniques such as principal components analysis (PCA) can still be used to realistically downsize the environmental space (e.g. Broennimann et al., 2011).

### 3. A case study: The climatic niche of the invasive African clawed frog

To illustrate the use of alpha-shapes for SDM, we used the case of an invasive species, the African clawed frog (*X. laevis*). This frog is native from sub-Saharan Africa, but its use in biological research and as a pet led to its introduction worldwide. Currently, *X. laevis* is established in parts of Asia, Europe, and North and South America (Measey et al., 2012). For the purposes of modeling, we collected the species distribution from FrogMAP (Underhill et al., 2013), which provided a total of 810 occurrence records along South Africa, Lesotho, and Swaziland. These records are provided at a quarter-degree square resolution and the large majority of them had been collected from 1996 to 2003. FrogMAP covers only part of the known distribution of *X. laevis*. We performed this omission intentionally as we also aimed to highlight the use of alpha-shapes for differentiating unoccupied environmental conditions that are sampled by the data from those that are not (i.e. non-analog conditions). For this purpose, we also recorded all grid cells that compose the spatial coverage of FrogMAP.

#### 3.1. Climatic data

As mentioned earlier, most profile techniques assume that each predictor is equally relevant to the delimitation of the species potential distribution. Thus, it is the researchers’ responsibility to ensure that this condition is met. Here, we used six macroclimatic gradients related to temperature and water availability, which, at the global scale, can be plausibly expected to have a determinant role on the distribution of *X. laevis*. Specifically, we used mean temperature of the coldest month, mean temperature of the warmest month, mean number of frost days, mean total annual precipitation, precipitation seasonality, and mean number of wet days. These predictors were calculated by using monthly climatic data available at the CRU v3.10.01 database (Harris et al., 2014) for the period 1974–2003. The data, originally at a half-degree square resolution, were downscaled to a quarter-degree square resolution by using thin-plate splines. Because of the high dimensionality of the predictors set, we used a PCA to synthesize the information into three main axes, which jointly explained 95.3% of the total variance in the data (Table A1 in Appendix A).

It is important to point out that in the absence of a data transformation technique that standardizes the environmental space (as in a PCA), it is advisable to rescale all predictors to the same interval if different units of measurement are being considered. This will allow a more intuitive and precise control of the value of  $\alpha$  (i.e. the radius of the circle or sphere that controls the level of adjustment of the alpha-shape to the species occurrences).

#### 3.2. Software tools

Here, we demonstrate how alpha-shapes can be implemented for SDM in R (R Development Core Team, 2013; version 3.0.2). We used the “alphashape3d” package version 1.0–4 (Lafarge et al., 2014) for computing 3D shapes, but 2D operations could be computed likewise. Finally, to work with spatial data, we used the “raster” package (Hijmans and van Etten, 2014). We provide our code with explanatory comments in Appendix B. It must be highlighted that this is just one of the modeling environments possible and that other computer applications implementing alpha-shapes can be found online.

#### 3.3. Model implementation and selection of $\alpha$

The computation of an alpha-shape is uniquely determined by the position of the species occurrences in the environmental space (i.e. their values for each predictor) and the value of the parameter  $\alpha$ . For a 3D space, starting from the convex hull of the point set ( $\alpha = \infty$ ), the alpha-shape gradually shrinks and starts to develop cavities as  $\alpha$  decreases until, for sufficiently small  $\alpha$ , the alpha-shape has zero

volume. Given the fact that a potentially large number of alpha-shapes exist (i.e. the alpha-spectrum), the challenge lies in selecting the one that most adequately represents the species' niche. Here, we have addressed and compared three distinct approaches that may be of use for future works applying alpha-shapes for SDM.

The first approach consists of selecting the alpha-shape having the maximum discrimination between presences and absences of the species from the alpha-spectrum. This criterion is considered here mainly to illustrate the potential risks of accounting for specificity (i.e. the proportion of absences correctly predicted) when predictions are made to new spatial or temporal domains. However, it may still be a useful choice for the purposes of interpolation. Here, we identified the alpha-shape with the maximum discrimination between presences and absences of *X. laevis*. Species' absence corresponded to the locations covered by FrogMAP, but in which no individual was recorded ( $n = 885$ ). Discrimination was measured by using the true skill statistic (TSS; Allouche et al., 2006). The TSS corresponds to the sum of specificity and sensitivity (i.e. the proportion of presences correctly predicted) minus 1. This statistic ranges from  $-1$  to  $1$ , where a value of  $1$  indicates perfect agreement while values of  $0$  or less indicate a performance no better than random (Allouche et al., 2006). This statistic was calculated for each value of  $\alpha$  starting from  $0$  and increasing at steps of  $0.001$ , until a consistent decrease in accuracy was observed (Fig. A1 in Appendix A).

The second approach consists of the identification of the smallest alpha-shape surrounding all occurrences and not having any hollow spaces (hereafter referred to as "minimum bounding envelope"; MBE). This bounding criterion is ecologically sound because, on the one hand, it reduces the chances of overestimating the species niche and, on the other hand, it classifies all inner climatic combinations as suitable, which is a plausible ecological assumption (Austin, 2002; Guisan et al., 2012; Webber et al., 2012; Whittaker, 1967). In simple words, the MBE is a geometric construct that tightly bounds the environmental conditions occupied by the species. Any looser adjustment than this implies performing some sort of extrapolation in the environmental space. This relates to the third approach, which concerns the use of expert-knowledge for the selection of the envelope. Expert-knowledge can be either the sole basis for the selection or can be used to refine the choice of  $\alpha$  after the MBE is identified. In some cases, the MBE may be too restrictive for what is considered to be the species fundamental niche (i.e. the set of environmental combinations suitable for the species). An example of this is when the envelope excludes environmental space between a few sparse occurrence records; however, there is an *a priori* knowledge that this low occupancy is caused by a limited survey effort and not because conditions are less suitable. In this case, a more general (i.e. closer to the convex-hull) bounding may be considered. Likewise, the generalization of the envelope may also be performed to accommodate experimental knowledge about the species' ecological tolerances. Regardless of the purpose, the shape and level of adjustment of any alpha-shape can be explored with relative ease by using 2D or 3D plots. In the case of the more complex 3D space, the "alphashape3d" package in combination with "rgl" (Adler and Murdoch, 2007) allows an interactive exploration of the envelope, including, among other possibilities, its rotation, zooming, and overlay with distribution records.

For our data, we started the identification of the MBE by testing increasingly higher values of  $\alpha$  (starting from  $0$  and increasing at steps of  $0.001$ ) until achieving an alpha-shape enclosing all species occurrences ( $\alpha = 0.292$ ; see R code in Appendix B for details). Next, we used interactive 3D plots to search for possible unfilled spaces in the shape (see examples in Fig. A2 in Appendix A). For our data, the MBE corresponds to the aforementioned alpha-shape ( $\alpha = 0.292$ ) because the 3D representation does not show any hollow. However, it should be emphasized that for other datasets, it may be necessary to further increase the value of  $\alpha$  in order to include any environmental space found in unfilled areas. For illustrative purposes, we also retained a generalized version of the MBE. In the absence of external information

supporting the choice, we focused on smoothing the more irregular transitions of the MBE boundary – which potentially increases their realism (Austin, 2002) – while still preserving the overall shape and size of this envelope, i.e. avoiding highly speculative extrapolations. For our data, we found that a value of  $\alpha = 0.7$  allowed a good balance between the two previous criteria.

### 3.4. Dealing with outliers and uncertainty

Alpha-shapes may be less affected by outliers than most other profile techniques because spurious records may only induce local modifications on the shape of the envelope. Nonetheless, the accommodation of extreme outliers in the geometric construct may lead to an unrealistic increase in the value of  $\alpha$ . As such, before proceeding with alpha-shape selection, it is advisable to search for potential outliers in the data. Similar to the selection procedure itself, this can be carried out by exploring the data by means of 2D or 3D plots. Here, we have used a carefully selected set of species occurrences and no outlier was apparent in our data (Fig. A3 in Appendix A).

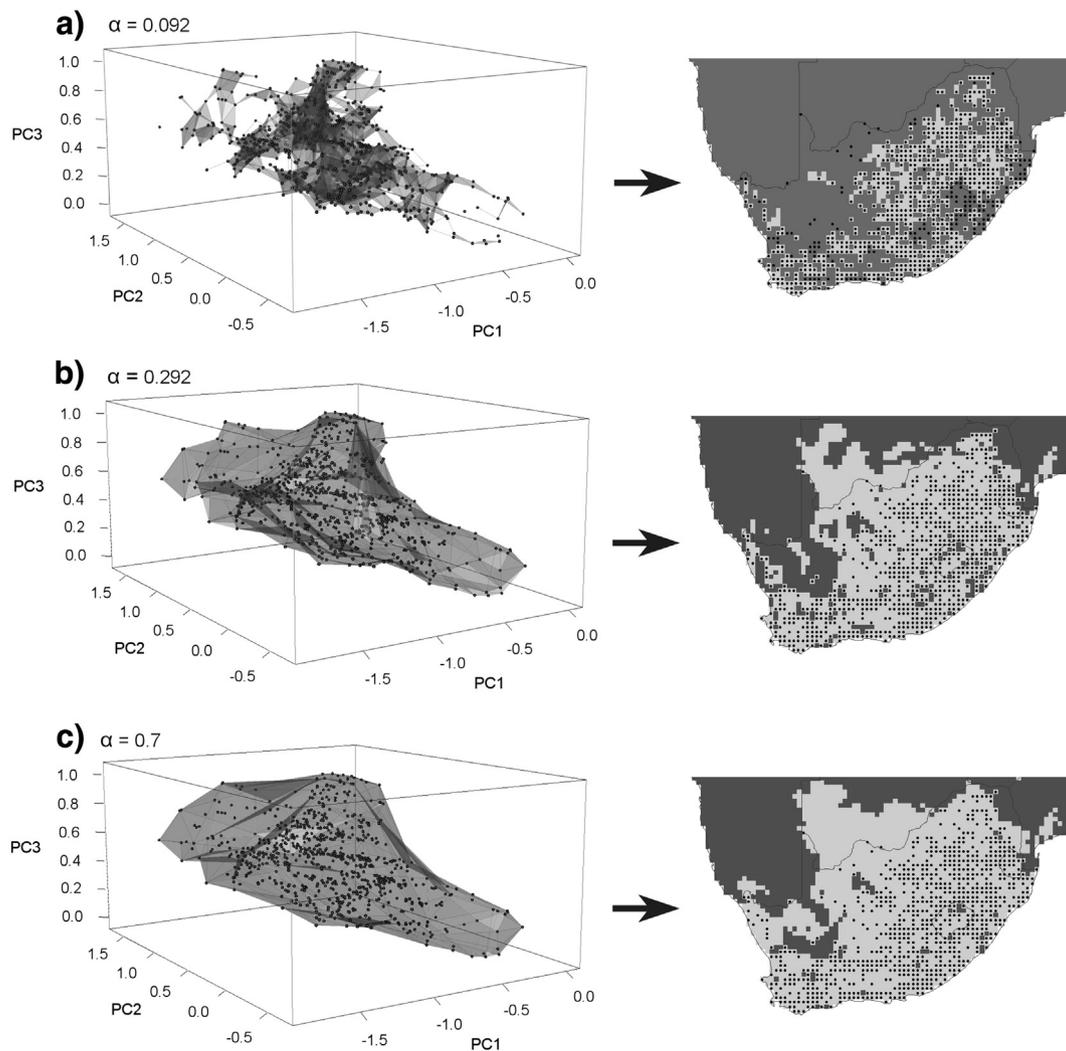
In addition, bootstrapping can also be valuable to assess the uncertainty of the predictions (e.g. Ponder et al., 2001). This procedure implies the generation of multiple predictions, each based on a random sample with replacement of the original occurrence dataset. The degree of overlap between the suitable areas predicted by each of the bootstrap-based models gives a measure of support for this classification. Low inter-model overlap indicates a sparsely occupied environmental space—as caused, for instance, by sink populations. Here, we evaluated the level of bootstrap support for the predictions based on the MBE. This evaluation was performed by calculating the level of agreement among 100 bootstrap replicates of this envelope (R code provided in Appendix B).

### 3.5. Delimiting non-analog conditions

In the context of SDM, non-analog conditions refer to environmental settings that are not sampled by the calibration data. Even in the absence of changes in the drivers of distribution not being considered in the models (e.g. biotic interactions or evolutionary adaptation in climate-based SDMs), predicting for these conditions is inherently uncertain because no information actually exists to base the predictions on (Fitzpatrick and Hargrove, 2009). As such, a recommended practice in SDMs extrapolating in space or time is to differentiate (e.g. on a map) the conditions that are indeed covered by the sample from those that are not (e.g. Capinha et al., 2012). This practice is conceptually and methodologically akin to the delimitation of the environmental envelope of a species. Therefore, alpha-shapes can also be a robust approach for delimiting non-analog conditions. Here, analog conditions refer to the full suite of climatic conditions covered by FrogMAP. We computed the minimum bounding envelope of this set of conditions ( $\alpha = 0.292$ ) and projected, across the globe, the climatic combination falling within (analogs) or outside (non-analogs) of it. We again recall that more generalized envelopes could be used (e.g. the convex-hull; García-López and Allué, 2013), but also emphasize that such generalizations should not be made arbitrarily and that their implications (i.e. the encompassing of unrepresented environmental settings) should be carefully assessed.

## 4. Result and discussion

We were able to obtain an alpha-shape with a high discrimination between the presence and absence of *X. laevis* (TSS =  $0.71$ ;  $\alpha = 0.092$ ). However, a visual examination of this envelope allowed verifying that some of the more peripheral occurrences were not included and also that diverse cavities exist in the midst of the point cloud (Fig. 2a). This is clearly a consequence of attempting to predict absences well. Most of *X. laevis* absences were positioned "in between" the



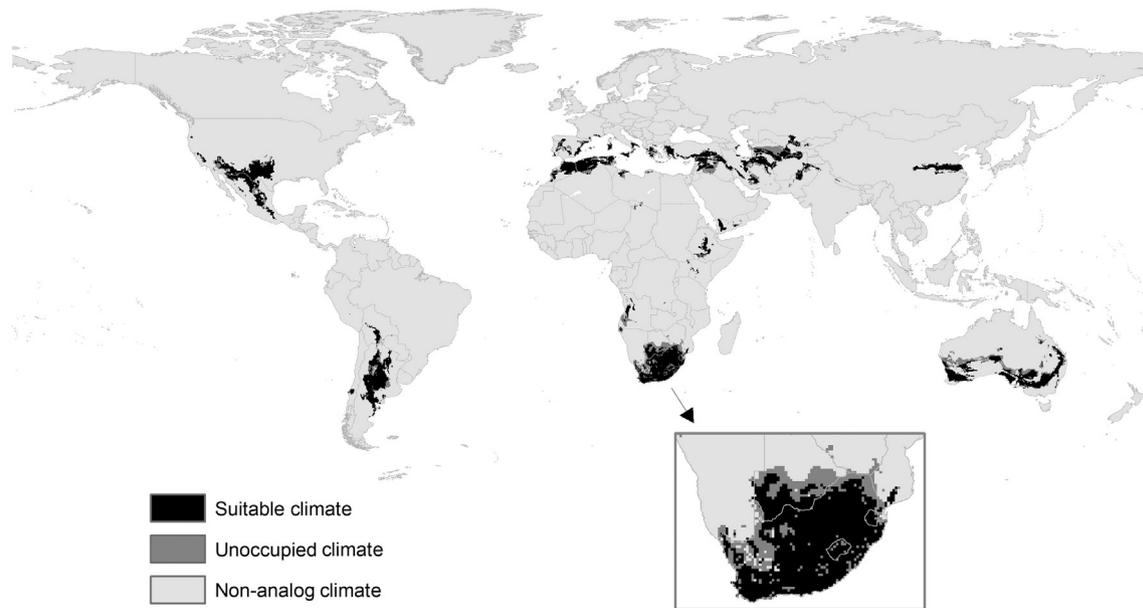
**Fig. 2.** Climatic suitability for *X. laevis* as predicted by three distinct alpha-shape envelopes: (a) the envelope having the maximum discrimination between the species' presences (black dots in the maps) and absences; (b) the smallest envelope without unfilled spaces that bounds all climatic conditions occupied by the species (i.e. the "minimum bounding envelope") and (c) an "expert-based" generalization of the previous.

occupied climatic space (Fig. A3 in Appendix A), clearly suggesting that they are caused by factors not being considered in the model (e.g. dispersal constraints or variability in local habitat availability). Predictably, the misguided attempt to exclude these absence records from the envelope – in order to increase specificity – also implied the subtraction of suitable climatic space and inherently the underestimation of the species' climatic tolerances. This is plainly visible in the mapped predictions, where a noteworthy number of species occurrences overlay climatic combinations that fall outside this "discrimination-optimized" envelope (Fig. 2a; black dots over dark gray areas). These results illustrate well the risks of accounting for specificity when models are intended for non-equilibrium settings (Jiménez-Valverde et al., 2008, 2011).

Alternatively, and as expected, both the minimum bounding envelope ( $\alpha = 0.292$ ; Fig. 2b) and its generalized version ( $\alpha = 0.7$ ; Fig. 2c) encompassed the full array of climatic conditions occupied by *X. laevis*. However, they differed on the level of extrapolation they make. While the MBE does not extrapolate in the climatic space – i.e. it does not predict beyond the conditions occupied by the species – the looser boundary of its generalized version leads to the encompassing of additional climatic space for which suitability is less certain. The level of uncertainty of this "extra layer" inherently depends on the reliability of the information supporting its inclusion. As our main objective here was to exemplify the generalization process, we replaced the use of solid

background information (as it would be desirable) by our intuitive perception of how the species niche could be like. Consequently, we assigned a high level of uncertainty to this additional "layer" of suitable climates and, as such, continued the interpretation of results based on the minimum bounding envelope.

In the absence of incorrect occurrence records or sink dynamics, the climatic conditions bounded by the MBE can be confidently classified as suitable for *X. laevis*. For instance, the climates in which *X. laevis* occurs in the study area are also found in the south central USA, Mexico, Bolivia, Argentina, Mediterranean basin, Middle East, East of China, and the South of Australia (Fig. 3; black areas). Accordingly, and in what concerns climate, we expect these areas to be vulnerable to *X. laevis* invasion. Furthermore, one can evaluate the reliability of this classification through the analysis of the bootstrap support values. These show that the large majority (77.5%) of the grid cells predicted as suitable by the MBE were also predicted by at least 75% of the bootstrap-based envelopes. Moreover, only 6.7% of the cells received less than 50% of support (Fig. A4 in Appendix A). These later areas – for which the uncertainty is high – are mainly found at the edges of the species potential range (Fig. A5 in Appendix A). Although low support values may be indicators of a detrimental effect of erroneous records or sink populations, it is important to bear in mind that these values only express a low frequency of the tested conditions in the species occurrence dataset. As such, other "legitimate" drivers of uncommonness – such as a reduced availability



**Fig. 3.** Worldwide prediction of the suitable climates for *X. laevis* based on the minimum bounding alpha-shape. Non-analog climates refer to climatic combinations that are outside the climatic space of the study area. Unoccupied climates refer to climatic combinations that are within the climatic space of the study area, but that are not occupied by the species.

of the tested conditions in the surveyed landscape – should be equally considered. In summary, the precise identification of classification errors would require further research.

Following the above-mentioned analyses, it is essential to consider the spatial extent of analog conditions. If the only coverage information available is the species occurrences, then the interpretation of the results is restricted to the areas being predicted as suitable (as above). Importantly, this does not mean that the remaining areas are unsuitable; instead, it is not possible to know if they meet the species requirements or not. Alternatively, if additional coverage information is available, the deduction of unsuitable areas becomes possible and occurs when the climatic space that was sampled is not fully encompassed by the species envelope. For instance, here, the sampled climatic space – given by the survey area of FrogMAP – goes beyond the conditions occupied by *X. laevis* (Fig. 3; dark-gray areas). A possible explanation is indeed that these conditions are outside the climatic tolerances of the species; however, other causes, such as dispersal barriers or detrimental biological factors (e.g. competitors or predators), should also be considered. Finally, another important outcome from our case study is that the climate space that was sampled corresponds to a limited portion of the climatic combinations existing worldwide (Fig. 3; light-gray areas). In other words, we simply cannot estimate the suitability of climates for *X. laevis* in most parts of the world.

## 5. Conclusions

Here, we have shown that alpha-shapes combine a particular set of features that can be highly valuable in the prediction of species distributions to new areas or time periods. This technique does not require absence or background data, makes no assumption about the shape of the species niche, and allows controlling the fit of the envelope to the environmental conditions captured by the occurrence data. Moreover, it allows – and to a certain extent, “forces” – modelers to have an influential (and expectedly informed) role in the modeling procedure. Conveniently, alpha-shapes can also be easily applied in the classification of non-analog conditions. We exemplified the capabilities of this technique by using a limited distribution dataset of *X. laevis* – a global invader. Using this information, we built a simple and intuitive, yet, ecologically sound representation of the species climatic niche and used it to identify

the 1) areas of the world matching the climates occupied in the surveyed area and 2) location of potentially unsuitable areas. We also showed that most parts of the world present climatic conditions that are not sampled by our data and for which climatic suitability for *X. laevis* is not possible to predict.

Despite the capabilities demonstrated earlier, it is not our intention to sustain the sole use of alpha-shapes in all spatial or temporal extrapolation exercises. For instance, if high-quality distribution data are available, it may be valuable to combine alpha-shapes with a probability density based model; the former can be used for delimiting suitable areas and non-analog conditions and the latter can be employed to differentiate between levels of potential occupancy. In addition, there is still room for improvement of alpha-shapes in the context of SDM. While having an environmental space that is constrained to up to three dimensions ensures interpretability, it also hinders model development when complex hyper-dimensional interactions are expected. As shown, workarounds are possible; however, the *de facto* implementation of alpha-shapes in higher dimensions would be a valuable accomplishment. Moreover, the possibility of using weighted alpha-shapes is also interesting (Edelsbrunner, 1992). This concept advances the construction of alpha-shapes to allow different levels of detail in the same envelope. This could be useful, for instance, to avoid the generalized degradation of detail that is caused by the enclosing of extreme records or sparsely occupied environmental space. Although computer applications exist for the construction of weighted alpha-shapes (e.g. Da et al., 2013), we anticipate that these are not yet sufficiently user-friendly for generalized use by ecologists.

In conclusion, we believe that the current practice of SDMs extrapolating in space or time has much to gain from alpha-shapes. This method combines a flexible envelope fitting—a characteristic often restricted to complex presence-absence models — with the simplicity and ecological interpretability of profile techniques. Moreover, its implementation in R is straightforward (see code in Appendix B) and the range of existing complementary analyses allows a detailed assessment of the results in both environmental and geographical spaces. Finally, we expect that this work will not only stimulate modelers to use alpha-shapes, but also help cultivate a general awareness about the importance of questioning the ecological realism of SDMs, particularly, when predictions are made for new spatial or temporal realms.

## Acknowledgment

C.C. was supported by a post-doctoral grant from the Portuguese Foundation for Science and Technology (FCT) (SFRH/BPD/84422/2012). We thank two anonymous reviewers for their constructive comments.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ecoinf.2014.06.001>.

## References

- Adler, D., Murdoch, D., 2007. rgl: 3D visualization device system (OpenGL). R package version 0.75. (URL <http://CRAN.R-project.org/package=rgl>).
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232.
- Austin, M., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157, 101–118.
- Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.* 3, 327–338.
- Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., Petitpierre, B., Pellissier, L., Yoccoz, N.G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N.E., Graham, C.H., Guisan, A., 2011. Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob. Ecol. Biogeogr.* 21, 481–497.
- Busby, J.R., 1991. Bioclim—a bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Melbourne, pp. 64–68.
- Calenge, C., Darmon, G., Basille, M., Loison, A., Jullien, J.M., 2008. The factorial decomposition of the Mahalanobis distances in habitat selection studies. *Ecology* 89, 555–566.
- Capinha, C., Leung, B., Anastácio, P., 2011. Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets. *Ecography* 34, 448–459.
- Capinha, C., Anastácio, P., Tenedório, J.A., 2012. Predicting the impact of climate change on the invasive decapods of the Iberian inland waters: an assessment of reliability. *Biol. Invasions* 14, 1737–1751.
- Carpenter, G., Gillison, A., Winter, J., 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 2, 667–680.
- Carroll, C., Zielinski, W.J., Noss, R.F., 1999. Using presence-absence data to build and test spatial habitat models for the Fisher in the Klamath Region, USA. *Conserv. Biol.* 13, 1344–1359.
- Cuevas, A., Fraiman, R., 2009. Set estimation. In: Kendall, W.S., Molchanov, I. (Eds.), *New Perspectives on Stochastic Geometry*. Oxford University Press, Oxford, pp. 374–397.
- Da, T.K.F., Loriot, S., Yvinec, M., 2013. 3D Alpha Shapes. In: CGAL Editorial Board (Ed.), *CGAL User and Reference Manual* (URL [http://doc.cgal.org/latest/Alpha\\_shapes\\_3/index.html](http://doc.cgal.org/latest/Alpha_shapes_3/index.html) [accessed 23 January 2014]).
- Diniz-Filho, J.A.F., Nabout, J.C., Bini, L.M., Loyola, R.D., Rangel, T.F., Nogueira-Bravo, D., Araújo, M.B., 2010. Ensemble forecasting shifts in climatically suitable areas for *Tropidacris cristata* (Orthoptera: Acridoidea: Romaleidae). *Insect Conserv. Divers.* 3, 213–221.
- Edelsbrunner, H., 1992. *Weighted alpha shapes*. Department of Computer Science, University of Illinois at Urbana-Champaign.
- Edelsbrunner, H., 2010. Alpha shapes—a survey. In: van de Weijgaert, R., Vegter, G., Icke, V., Ritzerveld, J. (Eds.), *Tessellations in the Sciences; Virtues, Techniques and Applications of Geometric Tilings*. Springer Verlag.
- Edelsbrunner, H., Mücke, E.P., 1994. Three-dimensional alpha shapes. *ACM Trans. Graph.* 13, 43–72.
- Edelsbrunner, H., Kirkpatrick, D., Seidel, R., 1983. On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* 29, 551–559.
- Eliith, J., Burgman, M., 2003. Habitat models for population viability analysis. In: Brigham, C.A., Schwartz, M.W. (Eds.), *Population Viability in Plants*. Springer-Verlag, New York, pp. 203–235.
- Eliith, J., Graham, C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32, 66–77.
- Eliith, J., Leathwick, J., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677–697.
- Eliith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Jacob, Peterson, T., Phillips, S., Richardson, K., Pereira, R., Schapire, R., Soberon, J., Williams, S., Wisz, M., Zimmermann, N., 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29, 129–151.
- Eliith, J., Kearney, M., Phillips, S., 2010. The art of modelling range-shifting species. *Methods Ecol. Evol.* 1, 330–342.
- Farber, O., Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecol. Model.* 160, 115–130.
- Fitzpatrick, M., Hargrove, W., 2009. The projection of species distribution models and the problem of non-analog climate. *Biodivers. Conserv.* 18, 2255–2261.
- García-López, J.M., Allué, C., 2013. Modelling future no-analogue climate distributions: a world-wide phytoclimatic niche-based survey. *Glob. Planet. Chang.* 101, 1–11.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.
- Guisan, A., Petitpierre, B., Broennimann, O., Kueffer, C., Randin, C., Daehler, C., 2012. Response to comment on “Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders”. *Science* 338, 193.
- Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int. J. Climatol.* 34, 623–642.
- Hijmans, R.J., van Etten, J., 2014. raster: geographic data analysis and modeling. R package version 2.2–5. (URL <http://CRAN.R-project.org/package=raster>).
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* 22, 415–442.
- Jiménez-Valverde, A., Lobo, J.M., Hortal, J., 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Divers. Distrib.* 14, 885–890.
- Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P., Lobo, J.M., 2011. Use of niche models in invasive species risk assessments. *Biol. Invasions* 13, 2785–2797.
- Lafarge, T., Pateiro-Lopez, B., Possolo, A., Dunkers, J.P., 2014. R implementation of a polyhedral approximation to a 3D set of points using the alpha-shape. *J. Stat. Softw.* 56, 1–19.
- Measey, G., Rödder, D., Green, S., Kobayashi, R., Lillo, F., Lobos, G., Rebelo, R., Thirion, J.-M., 2012. Ongoing invasions of the African clawed frog, *Xenopus laevis*: a global review. *Biol. Invasions* 14, 2255–2270.
- Munguia, M., Rahbek, C., Rangel, T.F., Diniz-Filho, J.A.F., Araújo, M.B., 2012. Equilibrium of global amphibian species distributions with climate. *PLoS One* 7, e34420.
- Nogueira-Bravo, D., 2009. Predicting the past distribution of species climatic niches. *Glob. Ecol. Biogeogr.* 18, 521–531.
- Pateiro-López, B., Rodríguez-Casal, A., 2013. Recovering the shape of a point cloud in the plane. *TEST* 22, 19–45.
- Peterson, A.T., 2011. Ecological niche conservatism: a time-structured review of evidence. *J. Biogeogr.* 38, 817–827.
- Ponder, W., Carter, G., Flemons, P., Chapman, R., 2001. Evaluation of museum collection data for use in biodiversity assessment. *Conserv. Biol.* 15, 648–657.
- 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.
- R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0. (URL <http://www.R-project.org/>).
- Rödder, D., Lötters, S., 2010. Explanative power of variables used in species distribution modelling: an issue of general model transferability or niche shift in the invasive Greenhouse frog (*Eleutherodactylus planirostris*). *Naturwissenschaften* 97, 781–796.
- Rotenberry, J.T., Preston, K.L., Knick, S.T., 2006. GIS-based niche modeling for mapping species' habitat. *Ecology* 87, 1458–1464.
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F., De Siqueira, M.F., Grainger, A., Hannah, L., 2004. Extinction risk from climate change. *Nature* 427, 145–148.
- Underhill, L.G., Navarro, R., Loftie-Eaton, M., 2013. Ongoing atlas of the amphibians in Southern Africa: the FrogMAP Virtual Museum. *FrogLog* 21, 54–56.
- Vaughan, I., Ormerod, S., 2005. The continuing challenges of testing species distribution models. *J. Appl. Ecol.* 42, 720–730.
- Walther, G., 1997. Granulometric smoothing. *Ann. Stat.* 25, 2273–2299.
- Walther, G., 1999. On a generalization of Blaschke's rolling theorem and the smoothing. *Math. Meth. Appl. Sci.* 22, 301–316.
- 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., 1967. Gradient analysis of vegetation. *Biol. Rev.* 42, 207–264.
- Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. *Pattern Recogn.* 37, 1–19.
- Zurell, D., Eliith, J., Schröder, B., 2012. Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Divers. Distrib.* 18, 628–634.