A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects

Mark P. Robertson, Martin H. Villet and Anthony R. Palmer

Department of Zoology and Entomology, Rhodes University, Grahamstown, 6140, South Africa.

Agricultural Research Council — Range and Forage Institute, PO Box 101, Grahamstown, 6140, South Africa

Abstract

A new predictive modelling technique called the fuzzy envelope model (FEM) is introduced. The technique can be used to predict potential distributions of organisms using presence-only locality records and a set of environmental predictor variables. FEM uses fuzzy logic to classify a set of predictor variable maps based on the values associated with presence records and combines the results to produce a potential distribution map for a target species. This technique represents several refinements of the envelope approach used in the BIOCLIM modelling package. These refinements are related to the way in which FEMs deal with uncertainty, the way in which this uncertainty is represented in the resultant potential distribution maps, and the way that these maps can be interpreted and applied. To illustrate its potential use in biogeographical studies, FEM was applied to predicting the potential distribution of three invasive alien plant species (*Lantana camara* L., *Ricinus communis* L. and *Solanum mauritianum* Scop.), and three native cicada species (*Capicada decora* Germar, *Platypleura deusta* Thun. and *P. capensis* L.) in South Africa, Lesotho and Swaziland. These models were quantitatively compared with models produced by means of the algorithm used in the BIOCLIM modelling package, which is referred to as a crisp envelope model (the CEM design). The average performance of models of the FEM design was consistently higher than those of the CEM design. There were significant differences in model performance among species but there was no significant interaction between model design and species. The average maximum kappa value ranged from 0.70 to 0.90 for FEM design and from 0.57 to 0.89 for the CEM design, which can be described as 'good' to 'excellent' using published ranges of agreement for the kappa statistic. This technique can be used to predict species' potential distributions that could be used for identifying regions at risk from invasion by alien species. These predictions could also be used in conservation planning in the case of native species.

Introduction

Biogeographical distribution models have been applied to a number of biological problems, and numerous examples can be found in recent reviews (Franklin, 1995; Guisan & Zimmermann, 2000). Most of these models can be classified as correlative, as they rely on strong, often indirect links between species distribution records and environmental predictor variables to make predictions (Beerling et al, 1995). Correlative models are an alternative to more complex mechanistic models that attempt to simulate the mechanisms considered to underlie the observed correlations with environmental attributes (Beerling et al, 1995; Robertson et al ., 2003).

Correlative distribution models can be divided into two groups based on the input data used to build them. Models that use both presence and absence locality records have been termed group discrimination techniques and those that use only presence locality records have been termed profile techniques (Caithness, 1995).

Presence/absence data are typically obtained by means of systematic field surveys (Margules & Austin, 1994; Austin, 1998) that are usually expensive and time consuming to conduct (Austin, 1998). As a result, presence-only data (obtained from museum or herbarium collections) are often used instead, despite their limitations (Funk & Richardson, 2002; Loiselle et al, 2003).

When presence-only data are used to make predictions, one can either use a profile technique or a group discrimination technique that relies on artificially generated pseudo-absence data (e.g. Lehmann et al, 2002; Zaniewski et al ., 2002).

A number of profile techniques have been described, examples include the algorithms used in BIOCLIM (Nix, 1986; Busby, 1991), DOMAIN (Carpenter et al ., 1993), those based on Principal Components Analysis (

Jones & Gladkov, 1999; Erasmus et al ., 2000; Robertson et al ., 2001) and factor analysis (Hirzel et al . 2002).

A number of group discrimination techniques have been described (reviewed by Guisan & Zimmermann, 2000). Currently, the most popular of these appear to be Generalized Linear Models (GLM; Austin et al, 1984; McCullagh & Nelder, 1989; Austin et al ., 1990; Austin et al ., 1994; Ferrier & Watson, 1997; Guisan et al ., 1998; Pearce & Ferrier, 2000) and Generalized Additive Models (GAM; Austin & Meyers, 1996; Ferrier & Watson, 1997; Hastie & Tibshirani, 1999; Leathwick & Whitehead, 2001).

An important question is whether one should use only presence data in a profile technique or make use of presence and artificially generated pseudo-absence data in a group discrimination technique. Although comparisons among techniques using these approaches have been undertaken (Ferrier & Watson, 1997; Loiselle et al, 2003) further work is needed to resolve this. An important finding is that certain techniques tend to make more conservative predictions than others (Loiselle et al, 2003) and thus contain fewer false positive prediction errors. An important factor in choosing among models is likely to be the cost of false negative vs. false positive errors in the predictions and on how the model's predictions are to be applied (Fielding & Bell, 1997; Loiselle et al, 2003). For example, in conservation planning, the cost of making a false positive prediction may be more costly than making a false negative prediction. It can be argued that when species range predictions are used for conservation area selection (e.g. Loiselle et al, 2003), it is better to have a conservative prediction (few false positives), which will ensure that the site selected to protect the species is suitable for that species rather than to select a site that is unsuitable. In contrast, when models are used to identify regions at risk from invasion by alien organisms, false negatives may be considered to be more costly. This is because sites with a high risk will be assigned low priorities, with the result that the species may invade the area and only be detected when it is well established.

In this paper we describe a simple profile technique (the fuzzy envelope model; FEM) that represents an alternative to using a group discrimination technique with pseudo-absence data.

Distribution modelling techniques have a conceptual base in Hutchinson's niche model (Schoener, 1990). Hutchinson's niche model consists of an abstract space defined by a set of axes, each of which represents a resource or condition of importance to the organism (Schoener, 1990). On each there will be a range of values within which the organism can survive. This space can be generalized mathematically to include as many axes as necessary to completely characterize the species' needs, resulting in an n-dimensional hyperspace that is termed the fundamental niche (Schoener, 1990). Few organisms occupy the whole of their fundamental niche because they may be excluded from parts of it by competition or predation (Begon et al, 1990). The reduced hyper volume in which the organism can survive is termed its realized niche (Begon et al, 1990; Schoener, 1990). The species environment relationship that forms the basis of correlative models is derived from a set of distribution records that are drawn from the target organism's realized niche (Austin, 2002).

The simplest of the profile techniques and some of the earliest models (Chicoine et al, 1985) are range-based models that are referred to as envelope models. These models are constructed as follows. A set of distribution records is used to derive values from each of a set of predictor variable maps (corresponding with each of the axes in Hutchinson's model). For each predictor variable, the minimum and maximum value is calculated from the values associated with the distribution records. Next, each predictor variable map is reclassified (using the minimum and maximum values) to produce a Boolean map, indicating regions of predicted presence (coded 1) and absence (coded 0) of the target organism. It is assumed that the minimum and maximum values (obtained from the training set) of each variable represent the climatic limits of the target organism (however, this may not be true for alien organisms that have not yet occupied all suitable sites). These maps are then superimposed using the Boolean AND function (Burrough, 1989), which is achieved by multiplying the Boolean maps (Hill & Binford, 2002). If the organism can survive somewhere in the map area, there will be a region where all of the 'survival ranges' overlap. In areas where not all conditions are satisfactory, the organism should be absent (Chicoine et al, 1985). The output consists of a binary potential distribution map indicating regions of predicted presence and absence. For the sake of clarity we refer to this approach as a simple envelope model (SEM).

A minor modification of the SEM has been implemented in a generic modelling package known as BIOCLIM (renamed ANUCLIM), which uses an envelope approach for predicting potential distributions of species (Nix, 1986; Busby, 1991). BIOCLIM has been used extensively to predict the potential distributions of various target organisms. It has been used to predict the potential distribution of various invasive plant species (Panetta &

Mitchell, 1991a,b; Sindel & Michael, 1992), various snakes (Nix, 1986), kangaroos (Skidmore et al ., 1996), gliders (Jackson & Claridge, 1999), the helmeted honeyeater (Pearce & Lindenmayer, 1998), the golden-tipped bat (Walton et al, 1992), Leadbeater's possum (Lindenmayer et al, 1991).
In BIOCLIM, the distribution of a target organism is predicted by characterizing the organism's tolerances in relation to a number of climatic parameters to produce a 'climate profile' for that organism (Busby, 1991). The parameters (predictor variables) are considered to provide a broad characterization of annual variations in temperature and levels of moisture availability (Nix, 1986). BIOCLIM predicts 'core' and 'marginal' environments for the organism under consideration, based on selected threshold values. Nix (1986) defined core environments as those values falling between the 5th and 95th percentile of each of the predictor variables, and marginal environments as those that fall outside of the core range but within the upper and lower limits of the variables, for the species. Core and marginal environments can be estimated using other reasonable values, e.g. Lindenmayer et al, (1991) used the 10th and 90th percentiles to define the core range. Nix (1986) points out that these thresholds are arbitrary. The output of BIOCLIM is a distribution map indicating regions of predicted presence (in terms of core and marginal habitat) and regions of predicted absence. Each of these regions is defined by classifying each of the localities in the map region into one of three Boolean or crisp sets (core, marginal or absent) based on the data in the training set. We refer to techniques that use this approach as crisp envelope models (CEM).

Fuzzy envelope model

The FEM represents a refinement of the CEM, which incorporates the notion that within a particular survival range, some conditions are more favourable than others and that the differences are continuous. This refinement uses a technique known as fuzzy classification, which is based on fuzzy set theory (Zadeh, 1965). Hill & Binford (2002) give a good introduction to fuzzy sets in the context of habitat models.

The use of fuzzy classification for potential distribution modelling has been investigated (Fairbanks & McKelly, 1994; Fairbanks, 1994), although it appears not to have seen much use in this field of biology (Chapman et al, 1994; Hill & Binford, 2002). Fuzzy classification has been used in soil science applications (Burrough, 1989; Lark & Bolam, 1997) and in remote sensing image classification techniques (Eastman, 1999).

Fuzzy classification is an approach that has its own algebra, which is an extension of Boolean algebra (Burrough, 1989). In classical mathematical set theory, an object either belongs to a particular set or not (Hill & Binford, 2002). These sets (classes) are typically characterized by a clearly defined value or criterion and are termed crisp sets (Lark & Bolam, 1997) or Boolean sets (Burrough, 1989). This type of classification assumes that all change between classes takes place at the class boundary and that very little significant change occurs within classes (Burrough, 1989), although this is often not the case with continuous data (Hill & Binford, 2002). In cases where a clearly defined criterion or value of class membership does not exist, fuzzy set theory can be used (Zadeh, 1965; Altman, 1994; Hill & Binford, 2002). A fuzzy set has a continuum of grades of membership, allowing for situations where clearly defined class membership values are absent. A fuzzy set is described by a fuzzy membership function, with values ranging from 0 to 1, corresponding with non-membership through to complete membership (Eastman, 1999). Fuzzy sets thus have continuous membership functions and for this reason the term 'continuous classification' is used by some authors instead of 'fuzzy classification' (Heuvelink & Burrough, 1993). Although fuzzy membership functions may appear to be similar to probability functions, these two concepts are quite different (Zadeh, 1965; Hill & Binford, 2002): fuzzy membership functions define possibility rather than probability (Zadeh, 1987).

In practice, a set of distribution records is used to derive a set of values from each of the predictor variable maps to produce a training data set. Each predictor variable map is reclassified using an appropriate fuzzy membership function. The shape of the membership function can be defined by the data in the training set. These reclassified maps (fuzzy sets) are then superimposed using fuzzy algebra (Heuvelink & Burrough, 1993) to produce a final map indicating the potential distribution of the target organism (a multivariate fuzzy set). The final potential distribution map contains a continuum of possibility values indicating conditions of varying suitability for the target organism. Localities with high possibility values are interpreted as representing more favourable conditions for the organism than those with low possibility values.

The FEM technique is explored by predicting the distribution of three invasive alien plant species and three cicada species in South Africa, Lesotho and Swaziland. In addition, an established and popular predictive modelling technique, the CEM, is used to quantitatively compare the performance of the FEM technique.

Methods

The data

The map region for this study included South Africa, Lesotho and Swaziland. Models were produced for three invasive alien plants, Lantana camara L, Ricinus communis L. and Solanum mauritianum Scop. and three cicada species Capicada decora Germar , Platypleura deusta Thun. and Platypleura capensis L. Alien plants were selected because they represent organisms for which absence data are usually not available or are unreliable, because they are of considerable economic and ecological importance and because we had some insight into the potential distribution of these species from previous work (Robertson et al ., 2001). Cicadas were selected because their datasets are representative of collections data, for which no absence data are available (Funk & Richardson, 2002). These species were selected because the datasets are quite small (each species is represented by fewer than 80 presence records), the species are endemic to the map region, and their distributions are moderately well understood.

The plant presence records were obtained from the Southern African Plant Invaders Atlas (Henderson, 1998), from the National Botanical Institute's PRECIS database, and by means of roadside surveys. Absence records were obtained by means of roadside surveys (Robertson et al, 2001) designed to sample major environmental gradients. The cicada presence records were obtained from the Albany Museum (Grahamstown), Durban Museum (Durban), National Museum of South Africa (Bloemfontein), National Collection of Insects (Pretoria), Natal Museum (Pietermaritzburg), Natural History Museum (London), Museum für Naturkunde (Berlin), Transvaal Museum (Pretoria), Rhodes University (Grahamstown), Pretoria University (Pretoria), and the private collections of Isak Coetzer, Rudi Mijburgh, Renzo Perissinotto, Martin Villet, Richard Stephen, Tony Ewart and published records of Michelle Boulard. The cicada absence records were based on localities selected using the expert opinion of one of the authors (MHV). For all of the species, only those distribution records that were recorded to the nearest minute were used for the purposes of modelling (to be consistent with the spatial resolution of the predictor variable data, see below). The numbers of presence and absence records available for each species for model training and testing are listed in Table 1. For each species, the set of presence records was partitioned randomly in a ratio of 3 : 1 training to testing data, and this was repeated five times to ensure that different combinations of records were available for model building and evaluation (Table 1). Localities representing the absence of these species were used for model evaluation but not for training (Table 1). There were thus five sets of data for model training (each set selected randomly from the available record set), consisting only of presence records. There were also five sets of data for evaluation, which consisted of presence and absence records.

Table 1 The number of presence (Pres.) and absence (Abs.) localities used for model training (Train.) and model evaluation (Eval.) of CEM and FEM models. For both the CEM and FEM techniques only localities representing the presence of the target species are used to train the models

	Train.	Eval.	
Species	Pres.	Pres.	Abs.
Lantana camara	322	64	46
Ricinus communis	237	47	30
Solanum mauritianum	324	65	33
Capcicada decora	27	5	6
Platypleura deusta	78	16	16
Platypleura capensis	23	5	7

Table 2 Predictor variables selected for building the distribution models

Monthly potential evaporation — January
Monthly potential evaporation — July
Monthly maximum temperature — January
Monthly minimum temperature — July
Monthly rainfall — January
Monthly rainfall — April
Monthly rainfall — July
Monthly rainfall — October
Number of days with frost
Altitude (Digital elevation model, one-minute resolution)

Digital maps of environmental variables (nine climatic variables and altitude) developed by Schulze et al. (1997) were selected as predictor variables (Table 2). Each of the climatic predictor variables was interpolated from point data obtained from a network of weather recording stations distributed throughout South Africa, to produce continuous digital maps at a one-minute spatial resolution (Schulze et al, 1997). The values of the 10 predictor variables (Table 2) associated with the training localities are referred to as the training data set.

Implementation

The crisp envelope model

The CEM algorithm was implemented in MATLAB (MATLAB, 2000) and the distribution maps were visualized in IDRISI32 (Eastman, 1999). Distribution maps containing regions indicating the core and marginal ranges for each species were produced. For a given species the following procedure was undertaken. Firstly, a map of the marginal range was produced. This map was generated by superimposing (by multiplication) a set of Boolean maps (one for each predictor variable). Each Boolean map was produced as follows. Each grid cell in the predictor variable map was evaluated and if its value fell within the range defined by the minimum and maximum value then it was assigned a value of one in the new Boolean map, otherwise it was assigned a value of zero. Next, a map of the core range was produced. This map was generated in the same way as the marginal range map except that the 10th and 90th percentiles were used to define the ranges for creating the Boolean maps from the predictor variable maps. Finally, the marginal and core range maps were superimposed (by addition) to produce a single map indicating the core and marginal range. We used the same approach described by Nix (1986) for BIOCLIM, except that we used different predictor variables.

The fuzzy envelope model

The design of the FEM was based on some of the concepts implemented in the fuzzy classification module of IDRISI32 (Eastman, 1999). The FEM program was written and implemented entirely in MATLAB. The FEM algorithm operates in the following way. For each predictor variable map, a fuzzily classified map is produced. These fuzzily classified maps are then superimposed using a minimum overlay function to produce the final fuzzy distribution map for the target organism. The minimum overlay function allows the individual fuzzy sets to be intersected to produce a multivariate set (Heuvelink & Burrough, 1993; Eastman, 1999; Hill & Binford, 2002) representing the potential distribution of the organism. Both Boolean and fuzzy multivariate sets are defined by a joint membership function (JMF), which describes the combined effect of the individual membership functions (Heuvelink & Burrough, 1993). The value of the JMF is given by the minimum value of the individual membership functions (Heuvelink & Burrough, 1993) and thus a minimum overlay function is appropriate (Eastman, 1999; Hill & Binford, 2002).

A fuzzy map is produced by assigning a value between zero and one to each grid cell in a predictor variable map using an appropriate fuzzy membership function. A number of forms of membership function could be used, such as linear, sigmoidal, j-shaped, or more complex non-monotonic shapes (Eastman, 1999). The FEM algorithm does not currently allow the data to directly define the shape of the membership function for each predictor variable, and this has to be specified by the user. Niche theory postulates that species' responses to environmental variables will at least be curvilinear and a bell-shaped (Gaussian) distribution is often assumed (Austin & Smith, 1989; Begon et al ., 1990), although skewed responses are common (Austin, 2002). A sigmoidal membership function was considered to be appropriate most often for approximating responses to environmental variables. The sigmoidal membership function can have symmetric, monotonically increasing or monotonically decreasing forms (Fig. 1; the equations defining these functions are listed in the Appendix).

A symmetric function is used to approximate a bell-shaped distribution that is often assumed to underlie resource utilization axes (Austin & Smith, 1989; Begon et al ., 1990). The use of monotonically increasing or monotonically decreasing functions allows for skewed distributions, which are considered to be the most common type of response function (Austin, 2002). The selection of the appropriate form of membership function (symmetric, monotonically increasing or monotonically decreasing), is done by examining a frequency histogram of the training data for that predictor variable.

The shape of the membership function is governed by four control points (Fig. 1), which are ordered from low to high on the measurement scale of the predictor variable axis. For the symmetric membership function, point 'a' marks the location where the membership function begins to rise above zero, point 'b' indicates where it reaches a value of one, point 'c' indicates the location where the membership grade begins to drop below one, and point 'd' marks the region where it again approaches zero (Fig. 1a). In the case of the monotonically increasing function, point 'a' indicates the location where the membership function rises above zero and all values on the measurement scale above point 'b' take a value of one. In the case of the monotonically decreasing function, all values below point 'c' on the measurement scale take on a value of one, while point 'c' indicates the location where the membership grade begins to drop below one.

For the symmetric membership function, points 'a' and 'd' are defined by the minimum and maximum values, respectively, and points 'b' and 'c' are both defined by the median value from the training data set. When the monotonically increasing function is selected, all values greater than or equal to the median are assigned a value of one. When the monotonically decreasing function is selected, all values less than or equal to the median are assigned a value of one.

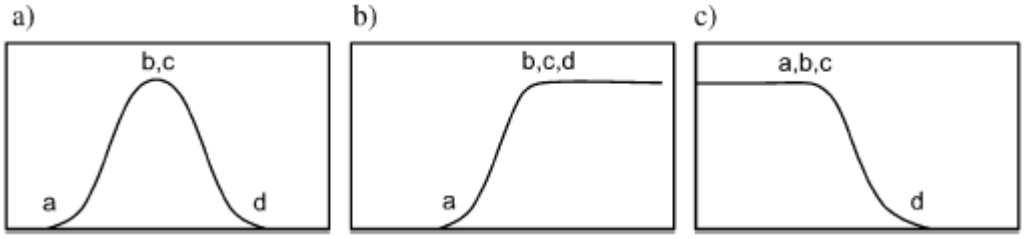

Figure 1 Sigmoidal fuzzy membership functions: (a) a symmetric membership function; (b) a monotonically increasing membership function and (c) a monotonically decreasing membership function (from Eastman, 1999). Control points are indicated by the letters a to d.

Figure 2 provides a summary of the implementation of the FEM technique and describes the process used to partition records into sets for training and evaluation.

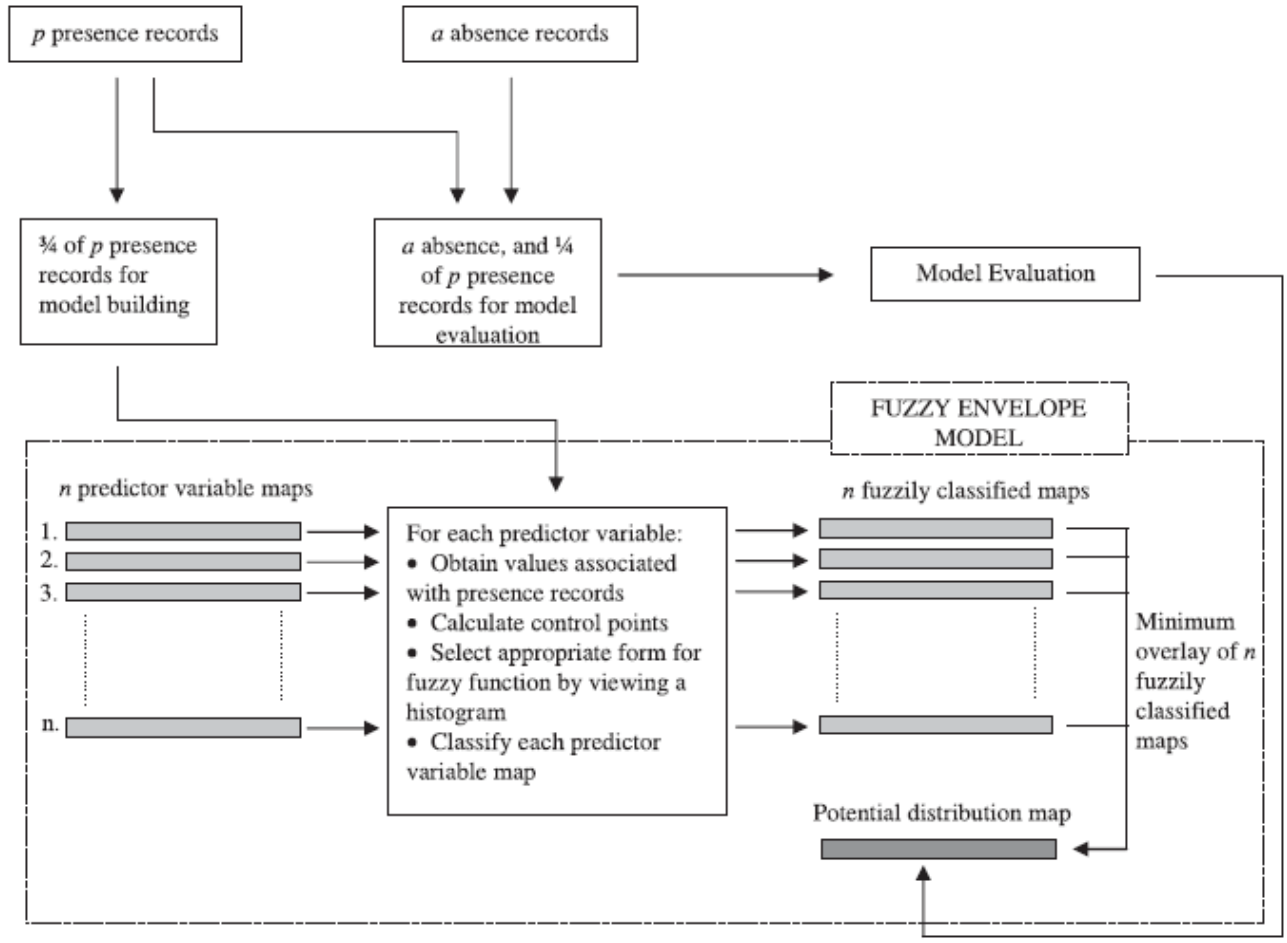

Figure 2 Implementation of the fuzzy envelope model (FEM) indicating how the original set of locality records was partitioned into a training and a testing (evaluation) set. The components of the FEM appear in the box. The procedure shown in this figure describes how a single potential distribution prediction is made. This procedure was repeated five times for each species so that different combinations of presence records could be used for model training and evaluation.

Model evaluation

Most quantitative model performance tests are based on a confusion matrix in which the observed (actual) and predicted presence/absence patterns are cross-tabulated (Fielding & Bell, 1997). Various threshold-dependent and threshold-independent accuracy measures can be calculated from the confusion matrix (reviewed by Fielding & Bell, 1997). Threshold-independent measures (e.g. ROC curves) are considered to be more robust and more objective than threshold-dependent measures (e.g. Kappa statistics) since they do not rely on a single threshold to distinguish between predicted presence and predicted absence (Fielding & Bell, 1997). However, threshold-independent accuracy measures could not be calculated for the CEM, since it does not produce a map of continuous values. As a result, the kappa statistic (a threshold-dependent measure) was calculated using those localities reserved for model evaluation (Table 1). In the case of the CEM design, the value of kappa was calculated at the threshold defined by the marginal range of the model. For FEM, kappa values were calculated at all thresholds between zero and one (instead of a single threshold) and the maximum value of kappa (κmax.) was selected as a measure of performance for the model. The threshold at which performance was highest was thus selected for each model, which can be regarded as an

optimum threshold for evaluating predictions on an independent data set (Guisan & Zimmermann, 2000). The equation for calculating the kappa statistic (Fielding & Bell, 1997) is:

$$\kappa = [(a + d) - (((a + c)(a + b) + (b + d)(c + d))/N]/ [N - (((a + c)(a + b) + (b + d)(c + d))/N)]$$

where:

a = the number of cases predicted present when actually present
b = the number of cases predicted present when actually absent
c = the number of cases predicted absent when actually present
d = the number of cases predicted absent when actually absent
N = a + b + c + d

Results

Results of a two-way ANOVA suggest that there was a significant difference in performance between models produced using the CEM and FEM designs (Table 3). The average performance of the FEM design was consistently higher than the CEM design. There were significant differences in model performance among species but no significant interaction between model design and species (Table 3).

Table 3 Results of a two-way ANOVA. Model performance is the mean of five maximum kappa values (based on five replicates) for each species

	d.f.	SS	MS	*F*	*P*-value
Model design	1	0.058	0.058	5.017	0.03
Species	5	0.411	0.082	7.136	0
Model * species	5	0.056	0.011	0.976	0.442
Residuals	48	0.553	0.012		

Potential distribution maps produced using the FEM and CEM techniques for all the species appear to correspond fairly well with localities used to build these models (Figs 3 and 4). This can be confirmed by the results of the quantitative tests of model performance. The average maximum kappa value ranged from 0.70 to 0.90 for FEM design and from 0.57 to 0.89 for the CEM design, which can be described as ‘good’ to ‘excellent’ using the ranges of agreement for the kappa statistic proposed by Monserud & Leemans (1992).

There is a difference in the appearance of the distribution map produced by the FEM design compared with the CEM design. Distribution maps produced from the FEM design have a continuous grade of values (which have been reclassified into five classes for better display) while those produced from the CEM have only two categories, the core and marginal range (Figs 3 and 4).

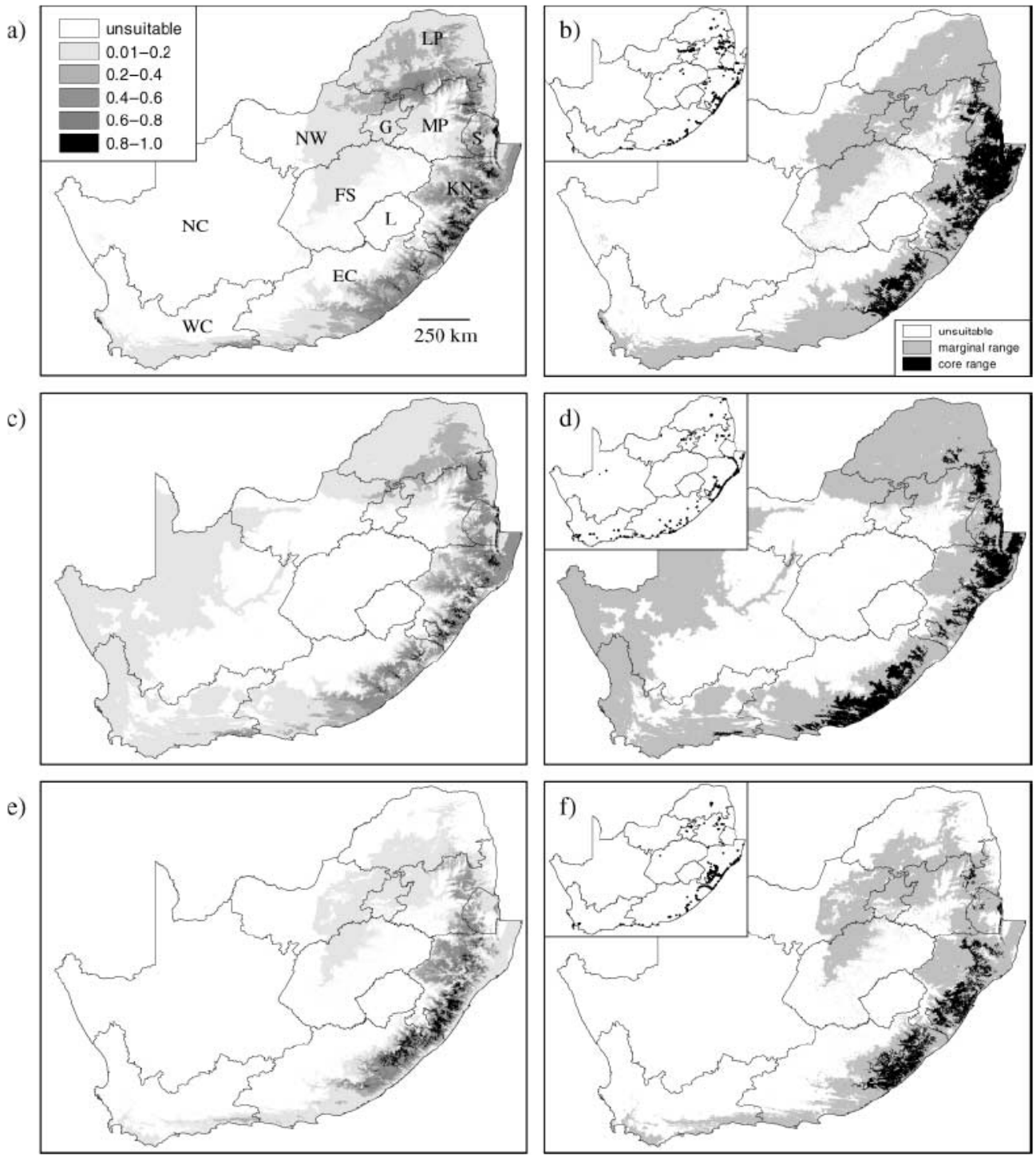

Figure 3 Potential distribution maps for three alien plant species generated using FEM and CEM predictive modelling techniques; a. *Lantana camara* FEM (κmax = 0.65, *n* = 322); Countries: Lesotho (L) and Swaziland (S); Provinces of South Africa: Limpopo (LP), Northwest (NW), Gauteng (G), Mpumalanga (MP), Northern Cape (NC), Free State (FS), KwaZulu-Natal (KN), Western Cape (WC) and Eastern Cape (EC);b. *Lantana camara* CEM (κmax = 0.65); c. *Ricinus communis* FEM (κmax = 0.87, *n* = 237); d. *R. communis* CEM (κmax = 0.58); e. *Solanum mauritianum* FEM (κmax = 0.74, *n* = 324); f. *S. mauritianum* CEM (κmax = 0.74). The insets indicate the presence localities for each species.

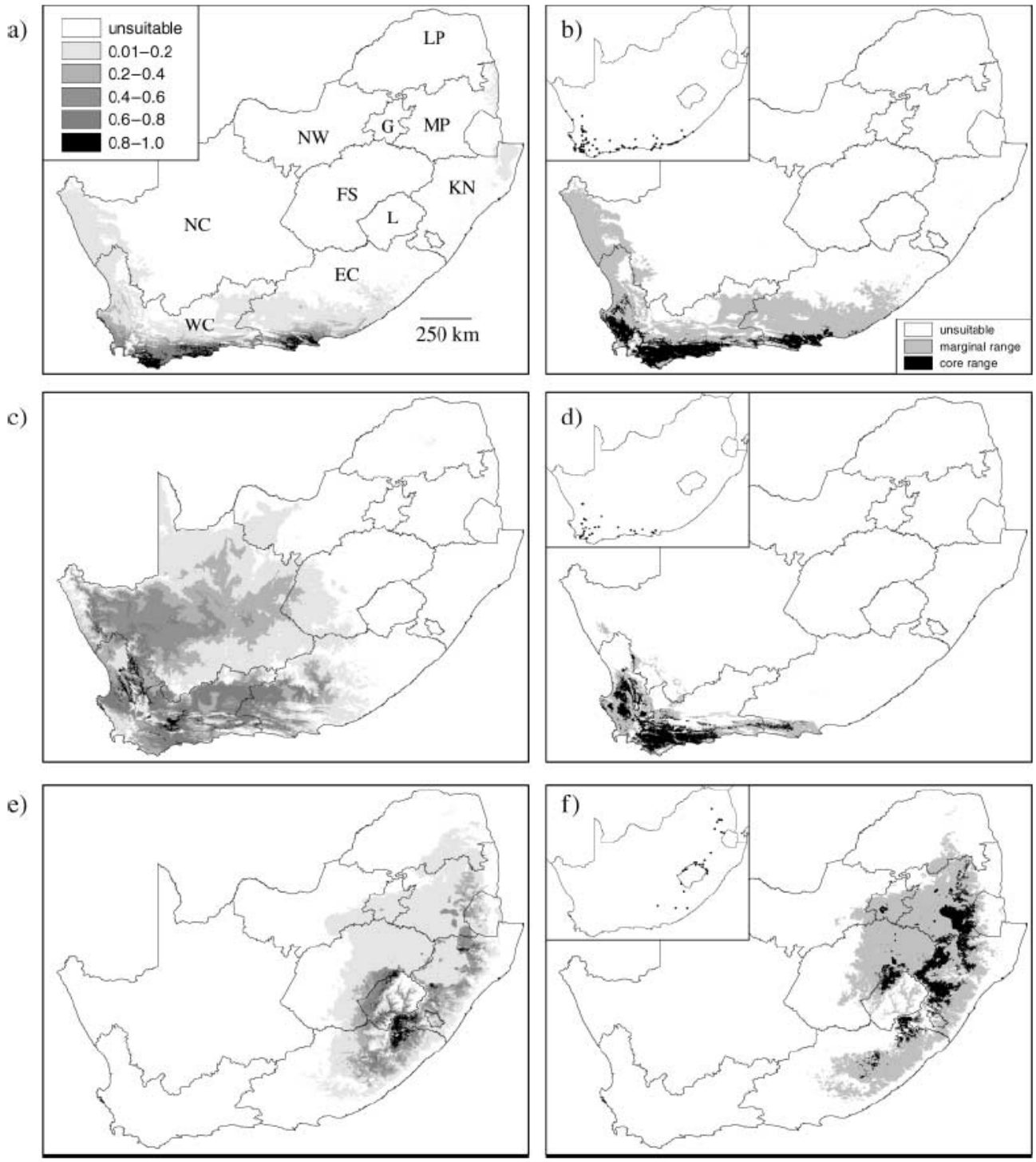

Figure 4 Potential distribution maps for three cicada species generated using FEM and CEM predictive modelling techniques; a. *Platypleura capensis* FEM (κmax = 0.94, *n* = 23); Countries: Lesotho (L) and Swaziland (S); Provinces of South Africa: Limpopo (LP), Northwest (NW), Gauteng (G), Mpumalanga (MP), Northern Cape (NC), Free State (FS), KwaZulu-Natal (KN), Western Cape (WC) and Eastern Cape (EC); b. *P. capensis* CEM (κmax = 0.88); c. *Capicada decora* FEM (κmax = 0.81, *n* = 27); d. *C. decora* CEM (κmax = 0.44); e. *Platypleura deusta* FEM (κmax = 0.82, *n* = 78); f. *P. deusta* CEM (κmax = 0.44). The insets Indicate the presence localities for each species.

There was very good visual agreement between maps generated from the FEM and CEM designs for R. communis and S. mauritianum (Fig. 3). The maximum kappa value (κmax.) was the same for the CEM and FEM designs for S. mauritianum, but the FEM design had a higher κmax value than the CEM model for R. communis. There was less visual agreement between the FEM and CEM designs for L. camara and P. deusta, and a considerable difference for C. decora (Fig. 4). The FEM predicted the presence of L. camara along the western border of the Limpopo Province, while the CEM predicted it to be absent in this region. The κmax value was the same for the FEM design (0.65) as for the CEM design (0.65). The CEM design appeared to make a much more conservative prediction than the FEM design in the case of C. decora. The κmax value was higher for the FEM design (0.81) than the CEM design (0.62), indicating better performance. For P. deusta, the FEM design predicted the species as being present in the southern and western regions of Lesotho, the northern regions of the Eastern Cape and in parts of the south-eastern Free State, whereas it was predicted as being absent in these regions by the CEM design. Again, the κmax value was higher for the FEM design (0.82) than the CEM design (0.44), indicating much better performance. In general, the core ranges of the CEM maps corresponded fairly well with areas of high possibility (> 0.6) in the FEM maps.

There was fairly good visual agreement between the distribution maps produced for L. camara, R. communis and S. mauritianum using the FEM technique (Fig. 3) and models produced using an independent PCA-based modelling technique (Robertson et al., 2001).

Discussion

Performance of FEMs

Models of the FEM design performed on average significantly better than those produced using the CEM design. For the selected distribution maps examined (Figs 3 and 4) there was either no difference in performance between the FEM and CEM designs or the FEM design performed better than the CEM design for models produced for the same species (see individual κmax values). These results suggest that the FEM design is capable of performing as well or better than the CEM design. In addition, the average performance of models of both the FEM and CEM designs was ‘good’ to ‘excellent’ using the ranges of agreement for the kappa statistic proposed by Monserud & Leemans (1992).

Design features of FEMs

One of the most important features of the design of FEMs is that various fuzzy membership functions are used to construct a model for a target species. Different forms of the membership function (symmetrical, monotonically increasing or monotonically decreasing) may be appropriate for expressing the response of the target species to a particular predictor variable. For example, a species that is intolerant of frost (which occurs quite frequently during winter in the higher elevation areas of the map region) would survive at localities where no frost occurs or where there are very few days of frost per year, but it would not survive at localities with moderate to large numbers of frost days per year.

A monotonically decreasing membership function would be appropriate to express this response. For this function, the possibility value is highest when the number of frost days is low (values less than the median) and decreases as the number of frost days increases (values greater than the median) to a minimum possibility value corresponding with the maximum value in the training set (d). A monotonically increasing membership function would be appropriate for expressing the response of a species that is associated with high altitudes. In this example, survivorship of the species increases with increasing altitude with no decrease in survivorship at very high altitudes. Plant species are found at the highest altitudes in the mountains of the map region; the highest point is 3515 m. For a monotonically increasing membership function, possibility values are lowest when the altitude is lowest (corresponding with the minimum value of the training set, a) increasing to a maximum at higher altitudes (corresponding with the median value of the training set, b).

A similar example of the response of a species to increasing altitude can be used to illustrate the use of a symmetric membership function. In this example, survivorship of the species increases with increasing altitude up to some maximum threshold, beyond which survivorship would again decline at very high altitudes.

A symmetric membership function would be appropriate to express this response as possibility values decrease for values larger than the median (c) to a minimum possibility value when the maximum training set value is reached (d).

The higher average performance of the FEM models over the CEM models is largely due to the greater flexibility of function selection of the FEM models and that it is possible to select an optimal threshold for the FEM but not for the CEM. When all of the functions selected in the FEM are symmetric then the ranges of an FEM and CEM model built on the same data will coincide, e.g. Fig. 3e,f. It is possible for FEMs to predict wider species ranges than CEMs. This occurs when a monotonically increasing or monotonically decreasing membership function is selected in the FEM for the predictor variable that is most limiting to the target species, e.g. Figs 4c,d. The ability to vary the threshold allows the predicted range of a FEM model to be reduced or expanded. The higher the suitability value at which the threshold is selected, the greater is the reduction in the predicted range. A reduction in the predicted range results in a more conservative prediction.

Fuzzy sets vs. crisp sets

Fuzzy sets (used in FEM) rather than crisp sets (used in CEM) may be more appropriate for building envelope distribution models, for various reasons. A locality with attributes that place it close to the threshold between two crisp classes (e.g. the coremarginal threshold) could be assigned to either class depending on the values of the thresholds in the training set. Heuvelink & Burrough (1993) suggest that it is not sensible to use crisp sets to classify continuous variables (the predictor variables) because attributes with very similar values may be assigned to different classes which have very different meaning, e.g. core and marginal.

It is often difficult or inappropriate to define attributes in terms of exact thresholds and when this is possible, there is often uncertainty as to the exact values of these thresholds (Heuvelink & Burrough, 1993; Hill & Binford, 2002). In such cases, fuzzy methods are appropriate because they are designed to handle this inexactness and uncertainty in a definable way (Burrough, 1989; Hill & Binford, 2002). In bioclimatic modelling this uncertainty arises because distribution models tend to predict the ‘average’ distribution of a species because climatic variables used as predictors are usually calculated using long-term means (Dent et al., 1989; Schulze et al., 1997). The locality records used to predict the potential distribution of an organism are usually collected over a period of time (usually several years or seasons). Inter-annual species range expansion or contraction may occur due to such factors as resource fluctuations or disturbances mediated by certain events, e.g. El Niño climate shifts (Hayward, 1997). This is likely to alter the values of the thresholds defining the distribution in the training set, depending on the temporal period during which the locality records were collected. Uncertainty also arises due to measurement and interpolation errors in the predictor variables (Heuvelink & Burrough, 1993). There is thus some uncertainty as to the exact value of these thresholds and hence the spatial extent of the core and marginal ranges of the target organism. We suggest that fuzzy classes more realistically represent this ‘average’ potential distribution and the uncertainty associated with the data than crisply defined techniques do. In addition, error propagation through models is reduced when continuous classes (fuzzy classes) are used rather than crisply defined classes (Heuvelink & Burrough, 1993).

Advantages of a continuous output

Through the use of fuzzy classification (fuzzy sets) a continuous output can be produced in the resultant distribution map. A continuous output allows one to calculate the kappa value for model evaluation at an optimal (e.g. where the kappa value is highest), rather than an arbitrary, threshold for that model. This is likely to be particularly important when comparing among model designs, species or sample sizes as it appears unlikely that a single arbitrary threshold has the same meaning among different model designs, species or sample sizes. It has been suggested that distribution maps produced from different modelling techniques (model designs) have different meanings (Zaniewski et al., 2002; Robertson et al, 2003).

Once the optimal threshold has been calculated (using the maximum kappa value) then the continuous distribution map can be objectively reclassified into a categorical distribution map containing only two classes (e.g. present and absent), which is often useful for further analysis and interpretation (Guisan & Zimmermann, 2000).

A continuous representation of the predicted distribution can be used in the final distribution map produced by the model to indicate to the user that there is a level of uncertainty in the prediction. The interpretation of fuzzy potential distribution maps by managers is likely to be different from the interpretation of binary potential distribution maps or maps with crisply defined core and marginal ranges. Fuzzy maps more realistically display the uncertainty associated with the input data used to generate them.

We suggest that the continuous output of the FEM design (and other techniques that produced continuous outputs) provides more scope for interpreting the predicted distribution of the target organism in terms of its biology than the output of the CEM design. Interpretation of the predicted distribution can be done in the following way. If one has data on the relative performance of the target species at a set of localities in the map region, such as density or fecundity, then the values in the model associated with those localities can be extracted. The relationship between the predicted values (extracted from the model) and the species performance measure can then be used to reclassify the continuous distribution map to produce a new map based on a feature of the biology of the target species. For example this approach was used to reclassify continuous maps of potential distribution predictions made for a number of biocontrol agents (insects) released in South Africa for the control of the invasive plant Lantana camara (Baars 2002). Data on the level of damage caused to the invader and the abundance of these agents at a number of localities were used as a measure of species performance for reclassifying the original continuous distribution maps into categorical maps to facilitate further analysis of the data. These data could also be used directly to define fuzzy membership functions, although this is not possible with the current version of the FEM. A further advantage of the FEM technique is that the individual fuzzily classified predictor variable maps that are used to produce the final distribution map can be examined and interpreted.

It is possible to refine the CEM approach, that uses only two envelopes (the marginal and core range), so that a continuous or nearly continuous output can be produced. This can be done by producing several successive envelopes, e.g. defined by the 5th and 95th, 10th and 90th, 15th and 85th percentiles. This would be equivalent to producing a fuzzy model using a stepped symmetrical function for each predictor variable.

Advantages and disadvantages of FEMs

We suggest that the major advantage of the FEM is that it does not rely on artificially generated pseudo-absence data. Alternative, more sophisticated, group discrimination techniques (such as GLM and GAM) require the use of pseudo-absence data in cases where only presence data are available (Lehmann et al., 2002; Zaniewski et al, 2002; Loiselle et al, 2003). Both the quality of the pseudo-absence records and the sensitivity of these techniques to false absences will influence the accuracy of the predictions. The quality of the records will largely depend on how the records are derived. One approach is to use randomly selected points as pseudo-absence records (Ferrier & Watson, 1997). A more reliable approach is to use the presence of one species as evidence for the absence of the target species (e.g. Zaniewski et al, 2002; Loiselle et al., 2003). This approach is only possible when a database of distribution records is available for a group of species in the same taxon (e.g. ferns, Zaniewski et al., 2002 or cotingid birds, Loiselle et al., 2003). A distinct advantage of GAMs is that the data directly define the shape of the response curves (Hastie & Tibshirani, 1999), whereas these have to be selected in a somewhat subjective manner by the user in the FEM. In addition, the shape of the response curves in GAM can be more complex than those currently available for FEM. Species responses are often better fitted using complex, nonparametric responses (available in GAM) than by linear or quadratic parametric curves (Bio et al., 1998). The availability of only three possible functions within FEM could be viewed as a weakness.

A criticism of BIOCLIM (which also holds for FEM) is that it does not account for interactions among predictor variables and each predictor variable axis is treated independently (Carpenter et al., 1993). Techniques that are based on multivariate statistics, for example, PCA (Erasmus et al., 2000; Robertson et al., 2001), GLM and GAM (Guisan & Zimmermann, 2000) take the multivariate structure of the data into consideration and may thus perform better. The CEM and FEM make the implicit assumption that all the predictor variables are equally important in predicting (or determining) the distribution of the target species. This is probably unrealistic and may lead to inaccuracies in predictions. In contrast, techniques such as GLM and GAM weight variables according to their importance and those that do not contribute significantly can be eliminated from the model.

In cases where pseudo-absence records can be generated with confidence (and the dataset contains few errors), models built using GLM or GAM may perform better than FEM. However, in cases where few data

points are available for a species, a simple approach such as FEM may be preferred to the data-driven approach used in GAM.

The role of correlative models in biological invasions

Recently, Hulme (2003) criticised the approaches that are currently utilized in monitoring and managing biological invasions by non-indigenous species. He raises some important points about mapping distributions and predicting species' ranges that are relevant here. Hulme (2003) points out that the spatial resolution and extent of species occurrence maps largely determines how they are interpreted and how they can be used to address issues in invasion ecology. While fine resolution mapping is considered to be most informative and most realistic (compared with coarse resolution maps) it demands considerable resources. Repeated sampling of specific points across a large geographical area is suggested to be more useful in terms of monitoring and management of invasions than attempting to produce maps of complete cartographic coverage of a species (Hulme, 2003). In addition, these points should be placed so that they sample a range of different environmental conditions, preferably using the gradsect approach (Austin & Heyligers, 1991). This will ensure that distribution data used to produce range predictions adequately sample the environment. While this may be the best approach to take in the future, much of the currently available data (for invasive alien and native species) is presence-only data. The approach described in this paper (FEM) will enable the use of currently available data, however, the limitations should be taken into account when applying the predictions.

Correlative distribution models assume that the species is in equilibrium with its environment (Austin, 2002; Hulme, 2003), however, this will not be true for species whose distributions are still expanding (many alien species). Hulme (2003) discusses the problems associated with predictions made for these species. Repeated sampling of specific points will help to distinguish those sites where the species is likely to be genuinely absent from those that are suitable but where the species has not yet invaded. For these species it will be necessary to iteratively refine distribution models as further data are gathered and as more insight is gained into the factors that are responsible for limiting their distributions. This will hopefully lead to models that incorporate factors that are more mechanistically important, leading to better predictions.

Hulme (2003) suggests that a wider range of variables should be used, and that variables such as land use, human population density and geology may be important. These variables could be incorporated to increase the value of range predictions. For example, a land cover map (derived from satellite imagery, e.g. Fairbanks et al., 2000) could be incorporated to identify transformed habitats, which may represent regions in the landscape that are likely to be invaded first. A similar approach has been used in conservation planning, where a land cover map was used to identify regions with high potential for ensuring the persistence of biodiversity (Wessels et al., 2000).

Conclusion

The FEM is suited to predicting distributions of species for which absence data are either not available, or are unreliable. The advantages of FEM are that it is easy to implement, simple and it does not rely on pseudo-absence data. Although criticisms can be levelled at the FEM, it appears to be useful and to produce reasonable results but it needs to be compared quantitatively with other competing techniques.

FEMs appear to deliver credible results and they represent refinements to the CEM approach used in the BIOCLIM modelling package. These refinements are related to the way in which FEMs deal with uncertainty, the way in which this uncertainty is represented in the resultant potential distribution maps, and the way that these maps can be interpreted and applied.

Acknowledgements

We thank the School of Bioresources Engineering and Environmental Hydrology (University of Natal), the Water Research Commission and the South African Country Study for Climate Change for the use of the climatic predictor variables; Lesley Henderson at the Southern African Plant Invaders Atlas for locality data; Craig Peter for collecting locality data, The National Botanical Institute for the use of data from the National Herbarium, Pretoria Computerized Information System (PRECIS). This work was funded by the National Department of Agriculture, Directorate of Agricultural Land Resource Management (previously the Directorate

of Resource Conservation) and by the National Research Foundation. We thank Antoine Guisan, Anthony Lehmann, and an anonymous referee for their comments on an earlier draft of this paper.

References

Altman, D. (1994) Fuzzy set theoretic approaches for handling imprecision in spatial analysis. International Journal of Geographical Information Systems, 8, 271–289.

Austin, M.P. (1998) An ecological perspective on biodiversity investigations: examples from Australian eucalypt forests. Annals of the Missouri Botanical Garden, 85, 2–17.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling, 157, 101–118.

Austin, M.P., Cunningham, R.B. & Fleming, P.M. (1984) New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. Vegetatio, 55, 11–27.

Austin, M.P. & Heyligers, P.C. (1991) New approaches to vegetation survey design: gradsect sampling. Nature conservation: cost effective biological surveys and data analysis (ed. by C.R. Margules and M.P. Austin), pp. 31–36. CSIRO, Canberra.

Austin, M.P. & Meyers, J.A. (1996) Current approaches to modelling the environmental niche of eucalypts: implications for management of forest biodiversity. Forest Ecology and Management, 85, 95–106.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. (1994) Determining species response functions to an environmental gradient by means of a beta function. Journal of Vegetation Science, 5, 215–228.

Austin, M.P., Nicholls, A.O. & Margules, C.R. (1990) Measurement of the realized niche, environmental niches of five Eucalyptus species. Ecological Monographs, 60, 161–177.

Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. Vegetatio, 83, 35–47.

Baars, J.-R. (2002) Biological control initiatives against. Lantana camara L. (Verbenaceae) in South Africa: an assessment of the present status of the programme, and an evaluation of Coelocephalapion camarae Kissinger (Coleoptera: Brentidae) and Falconia intermedia (Distant) (Hemiptera: Miridae), two new candidate natural enemies for release on the weed. PhD Thesis, Rhodes University, Grahamstown.

Beerling, D.J., Huntley, B. & Bailey, J.P. (1995) Climate and the distribution of Fallopia japonica: use of an introduced species to test the predictive capacity of response surfaces. Journal of Vegetation Science, 6, 269–282.

Begon, M., Harper, J.L. & Townsend, C.R. (1990) Ecology — individuals, populations and communities, 2nd edn. Blackwell Scientific Publications, Oxford.

Bio, A.M.F., Alkemade, R. & Barendregt, A. (1998) Determining alternative models for vegetation response analysis: a nonparametric approach. Journal of Vegetation Science, 9, 5–16.

Burrough, P.A. (1989) Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science, 40, 477– 492.

Busby, J.R. (1991) BIOCLIM — a bioclimatic analysis and prediction system. Nature conservation: cost effective biological surveys and data analysis (ed. by C.R. Margules and M.P. Austin), pp. 64–68. CSIRO, Melbourne.

Caithness, N. (1995) Pattern, process and the evolution of the African antelope (Mammalia: Bovidae). PhD Thesis, University of the Witwatersrand, Johannesburg.

Carpenter, G., Gillison, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation, 2, 667–680.

Chapman, A., Fairbanks, D.H.K. & Louw, J. (1994) Bioclimatic profiles for the potential afforestation of Pinus tecunumanii in the Eastern Transvaal. FOR-DEA 815, Division of Forest Science and Technology, CSIR.

Chicoine, T.K., Fay, P.K. & Nielsen, G.A. (1985) Predicting weed migration from soil and climate maps. Weed Science, 34, 57– 61.

Dent, M.C., Lynch, S.D. & Schulze, R.E. (1989) Mapping mean annual and other rainfall statistics in southern Africa, 1st edn. Department of Agricultural Engineering, Pietermaritzburg.

Eastman, J.R. (1999) Guide to GIS and image processing, Vol. 2. Clark Laboratories, Worcester.

Erasmus, B.F.N., Kshatriya, M., Mansell, M.W., Chown, S.L. & van Jaarsveld, A.S. (2000) A modelling approach to antlion (Neuroptera: Myreleontidae) distribution patterns. African Entomology, 8, 157–168.

Fairbanks, D.H.K. (1994) Report on the use of GIS modelling within FORESTEK for forest suitability mapping, and the possible further need for development of this technology. FOR-I 473. Division of Forest Science and Technology, CSIR.

Fairbanks, D.H.K. & McKelly, D. (1994) Investigating fuzzy set classification methods for use in GIS decision support modelling to determine land suitability. FOR-I 515. Division of Forest Science and Technology, CSIR.

Fairbanks, D.H.K., Thomson, M.W., Vink, D.E., Newby, T.S., Van den Berg, H.M. & Everard, D.A. (2000) The South African land-cover characteristics database: a synopsis of the landscape. South African Journal of Science, 96, 69–82.

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

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/ absence models. Environmental Conservation, 24, 38–49.

Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progress in Physical Geography, 19, 474–499.

Funk, V.A. & Richardson, K.S. (2002) Systematic data in biodiversity studies: use it or lose it. Systematic Biology, 51, 303–316.

Guisan, A., Theurillat, J.-P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an alpine environment. Journal of Vegetation Science, 9, 65–74.

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

Hastie, T.J. & Tibshirani, R. (1999) Generalized additive models. Chapman & Hall, London.

Hayward, T.L. (1997) Pacific ocean climate change: atmospheric forcing, ocean circulation and ecosystem response. Trends in Ecology and Evolution, 12, 150–154.

Henderson, L. (1998) Southern African plant invaders atlas (SAPIA). Applied Plant Sciences, 12, 31–32.

Heuvelink, G.B.M. & Burrough, P.A. (1993) Error propagation in cartographic modelling using Boolean logic and continuous classification. International Journal of Geographical Information System, 7, 231–246.

Hill, K.E. & Binford, M.W. (2002) The role of category definition in habitat models: practical and logical limitations of using Boolean, indexed, probabilistic and fuzzy categories. Predicting species occurrences: issues of accuracy and scale (ed. by J.M. Scott, P.J.
Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 97–106. Island Press, Washington, D.C.

Hirzel, A.H., Hausser, J., Chessel D. & Perrin, N. (2002) Ecological niche factor analysis: how to compute habitat suitability maps without absence data? Ecology, 83, 2027–2036.

Hulme, P.E. (2003) Biological invasions: winning the science battles but losing the conservation war? Oryx, 37, 178–193.

Jackson, S.M. & Claridge, A. (1999) Climatic modelling of the distribution of the mahogany glider (Petaurus gracilis), and the squirrel glider (P. norfolcensis). Australian Journal of Zoology, 47, 47–57.

Jones, P.G. & Gladkov, A. (1999) Floramap — a computer tool for predicting the distribution of plants and other organisms in the wild. International Center for Tropical Agriculture, Cali, Columbia.

Lark, R.M. & Bolam, H.C. (1997) Uncertainty in prediction and interpretation of spatially variable data on soils. Geoderma, 77, 263–282.

Leathwick, J.R. & Whitehead, D. (2001) Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. Functional Ecology, 15, 233–242.

Lehmann, A., Overton, J.M. & Leathwick, J.R. (2002) GRASP: Generalized regression analysis and spatial predictions. Ecological Modelling, 157, 187–205.

Lindenmayer, D.B., Nix, H.A., McMahon, J.P., Hutchinson, M.F. & Tanton, M.T. (1991) The conservation of Leadbeater's possum, Gymnobelideus leadbeateri (McCoy): a case study of the use of bioclimatic modelling. Journal of Biogeography, 18, 371–383.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T. & Smith, K.G. & Williams, P.H. (2003) Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology, 17, 1591–1600.

Margules, C.R. & Austin, M.P. (1994) Biological models for monitoring species decline: the construction and use of data bases. Philosophical Transactions of the Royal Society, London Series B, 344, 69–75.

MATLAB (2000) Using MATLAB (release 12), The MathWorks, Natick. McCullagh, P. & Nelder, J.A. (1989) Generalized linear models, 2nd edn. Chapman & Hall, London.

Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. Ecological Modelling, 62, 275–293.

Nix, H.A. (1986) A biogeographical analysis of Australian elapid snakes. Atlas of elapid snakes of Australia (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.

Panetta, F.D. & Mitchell, N.D. (1991a) Bioclimatic prediction of the potential distributions of some weed species prohibited entry to New Zealand. New Zealand Journal of Agricultural Research, 34, 341–350.

Panetta, F.D. & Mitchell, N.D. (1991b) Homoclime analysis and the prediction of weediness. Weed Research, 31, 273–284.

Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling, 128, 127–147.

Pearce, J. & Lindenmayer, D. (1998) Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (Lichenostomus melanops cassidix) in southeastern Australia. Restoration Ecology, 6, 238–243.

Robertson, M.P., Caithness, N. & Villet, M.H. (2001) A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. Diversity and Distributions, 7, 15–27.

Robertson, M.P., Peter, C., Villet, M.H. & Ripley, B.S. (2003) Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. Ecological Modelling, 164, 153–167.

Schoener, T.W. (1990) The ecological niche. Ecological concepts: the contribution of ecology to an understanding of the natural world (ed. by J.M. Cherrett), pp. 79–113. Blackwell Scientific Publications, Oxford.

Schulze, R.E., Maharaj, M., Lynch, S.D., Howe, B.J. & Melvil- Thomson, B. (1997) South African atlas of agrohydrology and climatology, 1st edn. Water Research Commission, Pretoria.

Sindel, B.M. & Michael, P.W. (1992) Spread and potential distribution of Senecio madagascariensis Poir. (fireweed) in Australia. Australian Journal of Ecology, 17, 21–26.

Skidmore, A.K., Gauld, A. & Walker, P. (1996) Classification of kangaroo habitat distribution using three GIS models. International Journal of Geographical Information Systems, 10, 441– 454.

Walton, D.W., Busby, J.R. & Woodside, D.P. (1992) Recorded and predicted distribution of the Golden-tipped Bat Phoniscus papuensis (Dobson, 1878) in Australia. Australian Zoologist, 28, 1–4.

Wessels, K.J., Reyers, B., Kruger, M. & van Jaarsveld, A.S. (2000) Incorporating land cover information into regional biodiversity assessments in South Africa. Animal Conservation, 3, 67–79.

Zadeh, L.A. (1965) Fuzzy sets. Information and Control, 8, 338– 353. Zadeh, L.A. (1987) Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and applications: selected papers by LA Zadeh (ed. by R.R. Yager, S. Ovchinnikov, R.M. Tong and H.T. Nguyen), pp. 193–218. John Wiley and Sons, New York.

Zaniewski, A.E., Lehmann, A. & Overton, J.McC. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling, 157, 261–280.

Appendix

The following equations describe the monotonically increasing and monotonically decreasing sigmoidal membership functions, from which the symmetric membership function is comprised. These equations are identical to those used in the fuzzy classification module in IDRISI32 (Eastman, 1999).

The monotonically increasing function:

$$\mu = \cos \alpha^2, \quad \text{and} \quad \alpha = (1 - (x - \text{point a})/(\text{point b} - \text{point a})) \times \pi/2$$

When $x >$ point b, $\mu = 1$.

μ = the fuzzy possibility value of a given grid cell in a particular predictor variable map

x = the value of the predictor variable in a given grid cell

point a and point b refer to control points a and b (Fig. 1), the minimum and median values, respectively, of the training set.

The monotonically decreasing function:

$$\mu = \cos \alpha^2, \quad \text{and} \quad \alpha = (x - \text{point c})/(\text{point d} - \text{point c}) \times \pi/2$$

When $x <$ point c, $\mu = 1$.

μ = the fuzzy possibility value of a given grid cell in a particular predictor variable map

x = the value of the predictor variable in a given grid cell

point c and point d refer to control points c and d (Fig. 1), the median and maximum values, respectively, of the training set.