Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques

Mark P. Robertson, Craig I. Peter , Martin H. Villet and Brad S. Ripley

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

School of Botany and Zoology, University of Natal, Pietermaritzburg 3209, South Africa

Botany Department, Rhodes University, Grahamstown 6140, South Africa

Abstract

Models used to predict species' potential distributions have been described as either correlative or mechanistic. We attempted to determine whether correlative models could perform as well as mechanistic models for predicting species potential distributions, using a case study. We compared potential distribution predictions made for a coastal dune plant (*Scaevola plumieri*) along the coast of South Africa, using a mechanistic model based on summer water balance (SWB), and two correlative models (a profile and a group discrimination technique). The profile technique was based on principal components analysis (PCA) and the group-discrimination technique was based on multiple logistic regression (LR). Kappa (κ) statistics were used to objectively assess model performance and model agreement. Model performance was calculated by measuring the levels of agreement (using κ) between a set of testing localities (distribution records not used for model building) and each of the model predictions. Using published interpretive guidelines for the kappa statistic, model performance was "excellent" for the SWB model ($\kappa=0.852$), perfect for the LR model ($\kappa=1.000$), and "very good" for the PCA model ($\kappa=0.721$). Model agreement was calculated by measuring the level of agreement between the mechanistic model and the two correlative models. There was "good" model agreement between the SWB and PCA models ($\kappa=0.679$) and "very good" agreement between the SWB and LR models ($\kappa=0.786$). The results suggest that correlative models can perform as well as or better than simple mechanistic models. The predictions generated from these three modelling designs are likely to generate different insights into the potential distribution and biology of the target organism and may be appropriate in different situations. The choice of model is likely to be influenced by the aims of the study, the biology of the target organism, the level of knowledge the target organism's biology, and data quality.

Author Keywords: Predictive biogeography; Mechanistic models; Correlative models; PCA; Logistic regression; *Scaevola plumieri*; Coastal dune plants

1. Introduction

Models for predicting species' potential distributions have been used in many fields of biology. Franklin (1995) and Guisan and Zimmermann (2000) have reviewed a number of these models and give examples of their application. A current question is which models perform best, given particular circumstances (Guisan and Zimmermann, 2000; Hirzel et al., 2001 and Zaniewski et al., 2002). We make use of a new opportunity to address this question empirically. We refer here to those models that make potential distribution predictions for a single taxon, rather than models used for predictive vegetation mapping, some of which are reviewed by Franklin (1995).

Predictive modelling techniques have been described as static or dynamic models (Beerling et al., 1995). Static models provide time-independent equilibrium predictions while dynamic models predict time-dependent dynamic responses to a changing environment (Beerling et al., 1995). Both types of model have in turn been divided into two groups, namely *correlative* and *mechanistic* techniques (Beerling et al., 1995). Correlative models rely on strong, often indirect, links between species distribution records and environmental predictor variables to make predictions (Beerling et al., 1995). These models use values of a predictor variable or more commonly a set of predictor variables associated with distribution records to classify the predictor variable or predictor variable hyperspace into presence–absence regions, suitability values or probabilities, which are visualised as maps. These predictor variables can be direct, resource or indirect gradients and tend to be distal rather than proximal. Austin (2002) defines proximal and distal as being to the position of the predictor in the chain of processes that link the predictor to its impact on the plant. Mechanistic models attempt to simulate the mechanisms considered to underlie the observed correlations with environmental attributes (Beerling et al., 1995) by using a detailed knowledge of the target species' physiological responses to environmental variables as well as life-history attributes (Stephenson, 1998). Such models have also been referred to as *ecophysiological models* (Stephenson, 1998) and *process orientated* models (Carpenter et al., 1993). In contrast to correlative models, mechanistic models do not use values of a predictor variable or predictor variables associated with distribution records to classify the predictor variable(s). The predictor variables used in mechanistic models tend to be resource or direct rather than indirect gradients. These predictor variables tend to be more proximal than those used in correlative models.

Stephenson (1998) maintains that the distinction between correlative and ecophysiological (mechanistic) models is often not clear. For example, he observes that ecophysiological studies of plants depend on empirical correlations to determine quantitative relationships between physiologically important factors and vegetation distribution. Similarly, correlative models have an ecophysiological basis when they employ predictor variables that are suspected to be of broad physiological importance to plants (Stephenson, 1998).

Predictive models have generally been used to predict the potential distribution of a target species under current climatic conditions or various climate change scenarios and to

determine the importance of selected climatic variables on the distribution of the target species. Recent interest in the possible consequences of global climate change has resulted in a number of studies focusing on the climatic controls of vegetation distribution (Stephenson, 1998). Mechanistic models are considered to be more promising at successfully predicting climatically induced changes in the distribution of plant species (Stephenson, 1998), as these models will be more robust under changed climatic conditions than correlative models as certain correlations may cease to apply under changed conditions (Prentice et al., 1992). While mechanistic models are likely to yield superior results to correlative models (particularly under climate change scenarios) they are often extremely time-consuming and more difficult to build, relying on a greater knowledge of the biology of the target organism than correlative models.

Correlative models are particularly suited to cases where an initial estimate of the potential distribution of an organism is required, especially when the biology of the organism is not well known. Correlative models also can be used to obtain some insight into factors that may be responsible for limiting the distribution of the target organism when its biology is not well known. These insights can then be used to incorporate mechanistically more important predictor variables into the model, thus making it more mechanistic. Through an iterative process, a greater understanding of the target organism's biology can be developed and further insights into the factors that limit its distribution may be obtained. This may culminate in developing a mechanistic distribution model. Stephenson (1998) suggests that correlative approaches may play an important role in understanding of the relationships between climate and species distributions by identifying potentially significant and previously overlooked physiological mechanisms.

Correlative models that use both presence and absence locality records to make predictions have been referred to as *group discrimination techniques*, while those that use only presence locality records have been referred to as *profile techniques* (Caithness, 1995). Examples of group-discrimination techniques include those models based on discriminant analysis (Rogers and Randolph, 1993; Rogers and Williams, 1993 and Rogers et al., 1996), Generalised Linear Models (Austin et al., 1984; Austin et al., 1990; Austin et al., 1994; Osborne and Tigar, 1992; Guisan et al., 1998; Higgins et al., 1999 and Manel et al., 1999; Cumming, 2000a and Cumming, 2000b), generalised additive models (GAM; Austin and Meyers, 1996; Leathwick et al., 1996; Leathwick, 1998; Pearce and Ferrier, 2000 and Leathwick and Whitehead, 2001) and decision-tree-based methods (Walker, 1990; Lees, 1994; Michaelsen et al., 1994 and Williams, 1994). Examples of profile techniques include models developed by Palmer and Van Staden (1992), Erasmus et al. (2000), Robertson et al. (2001), Hirzel et al. (2002) and the approaches used in the modelling packages known as BIOCLIM (Nix, 1986 and Busby, 1991) and DOMAIN (Carpenter et al., 1993). For examples of comparisons between group-discrimination and profile techniques, see Hirzel et al. (2001) and Zaniewski et al. (2002).

The choice between using a mechanistic or correlative approach will depend largely on the purposes of the study and the current state of knowledge of the biology of the target

organism. One of the central questions is whether correlative models can perform as well as mechanistic models for predicting species' potential distributions.

Most of the mechanistic models in the literature have tended to make predictions at the continental scale and usually for several species or functional types, e.g. Woodward and Williams (1987), Prentice et al. (1992) and Neilson (1995). However, Peter et al. (2003) recently developed a model based on water balance, temperature and plant phenology to predict the potential distribution of a single species (*Scaevola plumeri*) at a regional scale. The development of this water balance model presents an opportunity to compare quantitatively the success with which mechanistic and correlative approaches are able to predict the potential distribution of a target species.

In this paper we compare potential distribution predictions made using three static modelling approaches for a coastal dune plant (*S. plumieri*) along the coast of South Africa. We compare the performance of a mechanistic approach based on water balance (described in Peter et al., 2003) and two correlative models (a profile and a group discrimination technique). The profile technique is based on principal components analysis (PCA), and is described and implemented by Robertson et al. (2001). Similar profile techniques include the approach used in the FloraMap package (Jones and Gladkov, 1999), the approach used by Erasmus et al. (2000) and Ecological Niche Factor Analysis (ENFA; Hirzel et al., 2002).

The group-discrimination technique is based on multiple logistic regression (LR), a form of Generalised Linear Model (GLM; McCullagh and Nelder, 1983) that has been used frequently in biology (e.g. Austin et al., 1984; Austin et al., 1990; Austin et al., 1994; Nicholls, 1989; Leathwick and Mitchell, 1992; Osborne and Tigar, 1992; Austin and Meyers, 1996; Guisan et al., 1998; Guisan et al., 1999; Higgins et al., 1999 and Manel et al., 1999; Cumming, 2000a and Cumming, 2000b; Pearce and Ferrier, 2000 and Hirzel et al., 2001).

2. Materials and methods

2.1. The target species: *Scaevola plumieri*

S. plumieri represents a good test case for developing and comparing mechanistic and correlative models. Ecophysiological measurements could be made with relative ease since the plant is short, has broad leaves and stands of this plant were relatively easily accessible. Since it has an open, well ventilated canopy and it occurs on the same substrate (sand) throughout its range, certain simplifying assumptions could be made (Peter and Ripley, 2000). Making the necessary ecophysiological measurements that would be required to build a similar mechanistic model for other species (e.g. a large forest species), would be more challenging.

An important implicit assumption made by these models is that the target species is in equilibrium with its environment (in an ecological rather than a physiological sense), since these are static models (Guisan and Zimmermann, 2000 and Austin, 2002). Since *S.*

plumieri is indigenous to southern Africa, it is likely to be in equilibrium with its environment in the sense that it has had sufficient time to occupy all suitable sites (as opposed to an alien plant species that may not yet be in equilibrium with the environment). It is unlikely to compete directly with other species and thus is unlikely to be excluded from a particular site due to competition for space. Since it is confined to a narrow habitat, namely coastal sand dunes, we have been able to sample a large proportion of the plant's potential habitat along the South African coast. The chance of make false negative errors is low because the plant is conspicuous and thus is unlikely to have been recorded absent when it was present at a particular site. False positive errors are unlikely since *S. plumieri* is not easily confused with other species. This attribute of the plant has thus enabled accurate presence and absence locality data to be collected. This in turn has allowed reliable profile (PCA) and group-discrimination (LR) models to be built.

2.2. The data

The predictor variables used in these models consisted of various digital climatic variable maps and the response variable consisted of point distribution records for *S. plumieri*. The digital climatic variable maps were developed by Schulze et al. (1997) for South Africa, Lesotho and Swaziland. Each of these climatic maps was interpolated from point data obtained from a network of weather recording stations distributed throughout South Africa, to produce continuous digital maps at a 1-min spatial resolution (Schulze et al., 1997). Monthly maps of mean maximum temperature, mean minimum temperature, mean relative humidity and median rainfall (Schulze et al., 1997) were used in the water balance model (Peter et al., 2003) while further preprocessing of these maps was done to yield a smaller subset of variables that were used to develop the correlative models.

Localities where *S. plumieri* was found to be present or absent (Fig. 1) were obtained by direct surveys using a GPS, from herbarium specimens and from historical photographs (for details see Peter et al., 2003). Locality data were partitioned randomly into a set of training localities and a set of testing localities in a ratio of 3:1, based on Huberty's (1994) recommendations. The training localities consisted of 158 presence and 57 absence records and the testing localities consisted of 53 presence and 19 absence records. The training localities were used to calibrate the models and the testing localities were used for model evaluation. Locality data were used only to calibrate the PCA and LR models and not in the calibration phase of the water balance models.
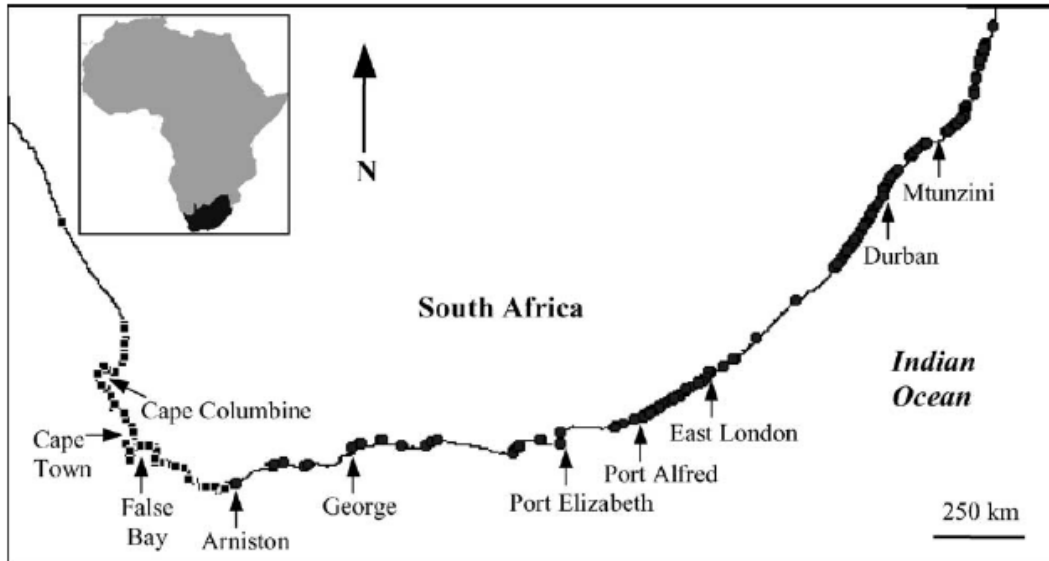
Fig. 1. The coast of South Africa indicating localities where *S. plumieri* was recorded as being present (•) and where it was recorded as being absent (▪). In the inset, black portion indicates the position of South Africa relative to Africa.

2.3. The water balance model

An empirical relationship of transpiration (E) to atmospheric vapor pressure deficit (VPD) was calculated at the leaf level for *S. plumieri* (Peter and Ripley, 2000). VPD can be calculated from atmospheric temperature and relative humidity. Peter and Ripley (2000) also successfully scaled leaf level transpiration rates to the canopy level. Transpiration rates of *S. plumieri* were extrapolated from VPD which was in turn calculated from regional level values of temperature and relative humidity. Water balance was calculated by subtracting transpiration from rainfall for a given month (Peter et al., 2003). Monthly maps of mean maximum temperature, mean minimum temperature, mean relative humidity and median rainfall were used to calculate 12 monthly water balance maps using the approach and equations described by Peter et al. (2003). These monthly water balance maps were cross-correlated to investigate their relationships (Table 1). Based on these correlations maps for October, November, December, January, February and March were summed to produce a map of summer water balance (SWB). Similarly maps for May, June, July and August were summed to produce a map of winter water balance (WWB). Seasonal, rather than annual, water balance was calculated because *S. plumieri* was found at sites that experienced water surpluses during the summer months, when the plants were most actively growing and reproducing (Peter et al., 2003).

Table 1. Correlation matrix of monthly water balance values for the entire coast ($n=1439$)

Month	January	February	March	April	May	June	July	August	September	October	November
February	0.992[a]										
March	0.979[a]	0.981[a]									
April	0.830[a]	0.839[a]	0.862[a]								
May	0.216[a]	0.208[a]	0.269[a]	0.676[a]							
June	−0.200[a]	−0.210[a]	−0.160[a]	0.323[a]	0.900[a]						
July	−0.190[a]	−0.200[a]	−0.150[a]	0.326[a]	0.901[a]	0.989[a]					
August	0.039	0.019	0.088[a]	0.502[a]	0.958[a]	0.939[a]	0.948[a]				
September	0.735[a]	0.723[a]	0.783[a]	0.915[a]	0.762[a]	0.440[a]	0.449[a]	0.661[a]			
October	0.927[a]	0.921[a]	0.959[a]	0.906[a]	0.464[a]	0.056[a]	0.068[a]	0.317[a]	0.909[a]		
November	0.986[a]	0.984[a]	0.991[a]	0.861[a]	0.283[a]	−0.140[a]	−0.130[a]	0.113[a]	0.796[a]	0.965[a]	
December	0.996[a]	0.987[a]	0.980[a]	0.829[a]	0.230[a]	−0.180[a]	−0.170[a]	0.064[a]	0.756[a]	0.939[a]	0.990[a]

[a] Significant correlations ($P < 0.05$).

We refer to the SWB model as a *mechanistic* model (although this term was not used to describe this model by Peter et al., 2003). We believe that SWB model is mechanistic for two reasons. Firstly, a predictor variable (SWB) was calculated using the physiological responses of the plant to environmental variables obtained by means of field measurements and through a knowledge of life history attributes (phenology). Secondly, the predictor variable was not classified using distribution records to produce a distribution map. This was unnecessary because the values of the predictor variable (SWB) had direct physiological significance to the plant. The plant should be absent from those sites experiencing summer water deficits and present at sites experiencing summer water surpluses (Peter et al., 2003). This is based on the hypothesis that the plant is unable to survive at sites that experience water deficits during those periods when the plant was actively growing and reproducing (i.e. summer).

2.4. Predictor variable pre-processing

To reduce the dimensionality of available climatic variable data, four PCAs were performed on the 12 maps for each of the following: mean monthly maximum temperature, mean monthly minimum temperature, mean monthly relative humidity and median monthly rainfall. Each of the monthly maps was calculated from mean values for a calendar month, e.g. January. PCA has previously been employed as a pre-analytical data reduction technique used in distribution modelling (Osborne and Tigar, 1992; Buckland and Elston, 1993; Robinson et al., 1997 and Guisan et al., 1998). Those principal component axes whose eigenvalues were greater in magnitude than eigenvalues obtained from datasets of random numbers of the same sample size were retained as predictor variables. This follows the "broken stick" stopping rule for PCA (Jackson, 1993). Predictor variables used in the PCA and logistic regression models are listed in Table 2.

Table 2. Predictor variables and their abbreviations used in the PCA and LR models

Abbreviation	Predictor variable
RH1	Component axis 1 of a PCA on 12 monthly mean relative humidity maps
RH2	Component axis 2 of a PCA on 12 monthly mean relative humidity maps
MXT1	Component axis 1 of a PCA on 12 monthly maximum temperature maps
MXT2	Component axis 2 of a PCA on 12 monthly maximum temperature maps
MNT1	Component axis 1 of a PCA on 12 monthly minimum temperature maps
MNT2	Component axis 2 of a PCA on 12 monthly minimum temperature maps
RN1	Component axis 1 of a PCA on 12 monthly rainfall maps
RN2	Component axis 2 of a PCA on 12 monthly rainfall maps

Although variables such as potential evaporation or elevation (as used previously, Robertson et al., 2001) could have been used in addition to those predictor variables listed for building the correlative models (Table 2), the same sets of predictor variables were used to build the mechanistic model and the two correlative models so that any differences in model performance could be attributed to model design and were not confounded by differences in the predictor variables used.

2.5. The PCA model

We used a PCA-based modelling technique, described by Robertson et al. (2001), for predicting environmental suitability for a target organism from environmental predictor variables using only presence locality records. This technique constructs a hyperspace for the target species using principal component axes derived from a training data set. The training data set comprises the values of the predictor variables associated with those localities where the species has been recorded as present. The origin of this hyperspace is taken to characterise the centre of the niche of the organism. All the localities (grid-cells) in the map region are then fitted into this hyperspace using the values of the predictor variables at these localities (termed the prediction data set). The Euclidean distance from any locality to the origin of the hyperspace gives a measure of the "centrality" of that locality in the hyperspace. These distances are used to derive probability values for each grid cell in the map region. The probability values are taken to indicate the suitability of each grid cell in a map for the target species in terms of the suite of predictor variables. The approach taken in this paper is identical to that of Robertson et al. (2001) except that different predictor variables were used.

2.6. The logistic regression model

A logistic regression was performed using GLMFIT, a generalised linear modelling function within MATLAB. In order to calculate the values of the coefficients (β_i in Eq. (1)), a binomial error distribution and a logit link function were used. The response variable consisted of localities representing surveyed presence (coded 1) and surveyed absence (coded 0). The response variables consisted of the values of the environmental variables associated with the surveyed localities. Probability values for each grid cell in

the map region were calculated by substituting the values of the predictor variables associated with that cell into the following equations:

$$\beta_1 variable_1 + \beta_2 variable_2 + \ldots + \beta_8 variable_8 = \eta \tag{1}$$

$$P_{(y=1)} = \frac{\exp(\eta)}{1 + \exp(\eta)} \tag{2}$$

The first equation is known as the linear predictor and the second equation is the inverse logistic transformation. In order to constrain the values of the linear predictor between 0 and 1, the inverse logistic transformation has to be applied.

2.7. Model evaluation

There are several measures for assessing model performance (Fielding and Bell, 1997 and Guisan and Zimmermann, 2000). A number of these measures are derived from a confusion matrix (Table 3). A reliable and well-known measure based on the confusion matrix is the kappa (κ) statistic (Fielding and Bell, 1997). The κ statistic is dependent on a single threshold to distinguish between predicted presence and predicted absence and thus falls into the class of threshold-dependent measures (Fielding and Bell, 1997). Threshold-independent measures, such as receiver operating characteristic (ROC) plots which are emerging as useful measures of model performance (Packer et al., 1999; Cumming, 2000a and Cumming, 2000b; Robertson et al., 2001) are considered to be superior since they use a range of thresholds and are therefore less likely to introduce distortions (Fielding and Bell, 1997). Although the ROC measure is considered to be a superior measure, it could not be used on our data since it requires values to be constrained between 0 and 1, and in the case of the water balance models, negative values were evident. As a result, we used the κ statistic to evaluate the models.

Table 3. A confusion matrix used to calculate κ statistics (Fielding and Bell, 1997)

Predicted	Observed	
	+	−
+	a	b
−	c	d

+: Presence; −: absence. The parameters a, b, c and d represent counts rather than percentages.

Kappa statistics were calculated from the parameters in the confusion matrix (Table 3) in two ways. Firstly, κ-values were calculated for all models using the presence and absence testing localities as *observed presence* and *observed absence*, respectively (Table 3). This gives a measure of *model performance*. Secondly, κ-values were calculated using the SWB model. Grid-cells with water balance values above zero were taken to represent

observed presence and grid-cells with water balance values below zero were taken to represent *observed absence*. This gives a measure of *model agreement* between the SWB model and the PCA model, and between the SWB model and the LR model. Kappa statistics were calculated using thresholds that yielded maximum κ-values for the PCA and LR models, following recommendations of Guisan and Zimmermann (2000). In contrast, κ-values for summer and winter water balances were calculated using thresholds of zero, which represent a biologically justifiable threshold. *S. plumieri* is unlikely to be able to survive water deficits (values below zero) for extended periods, particularly in the summer when it is actively growing (see Peter et al., 2003).

3. Results

S. plumieri was recorded as being present along the south and east coasts of South Africa (Fig. 1 and Fig. 2). Arniston was the most westward locality at which *S. plumieri* was recorded present. All localities west of Arniston represent observed absence and all localities to the east represent observed presence (Fig. 1 and Fig. 2). For simplicity, the coast to the north of Cape Columbine is referred to as *west coast*, between Cape Columbine and Arniston as the *south west coast*; between Arniston and Port Elizabeth as the *south coast* and between Port Elizabeth and the Mozambique border as the *east coast* (Fig. 2a).
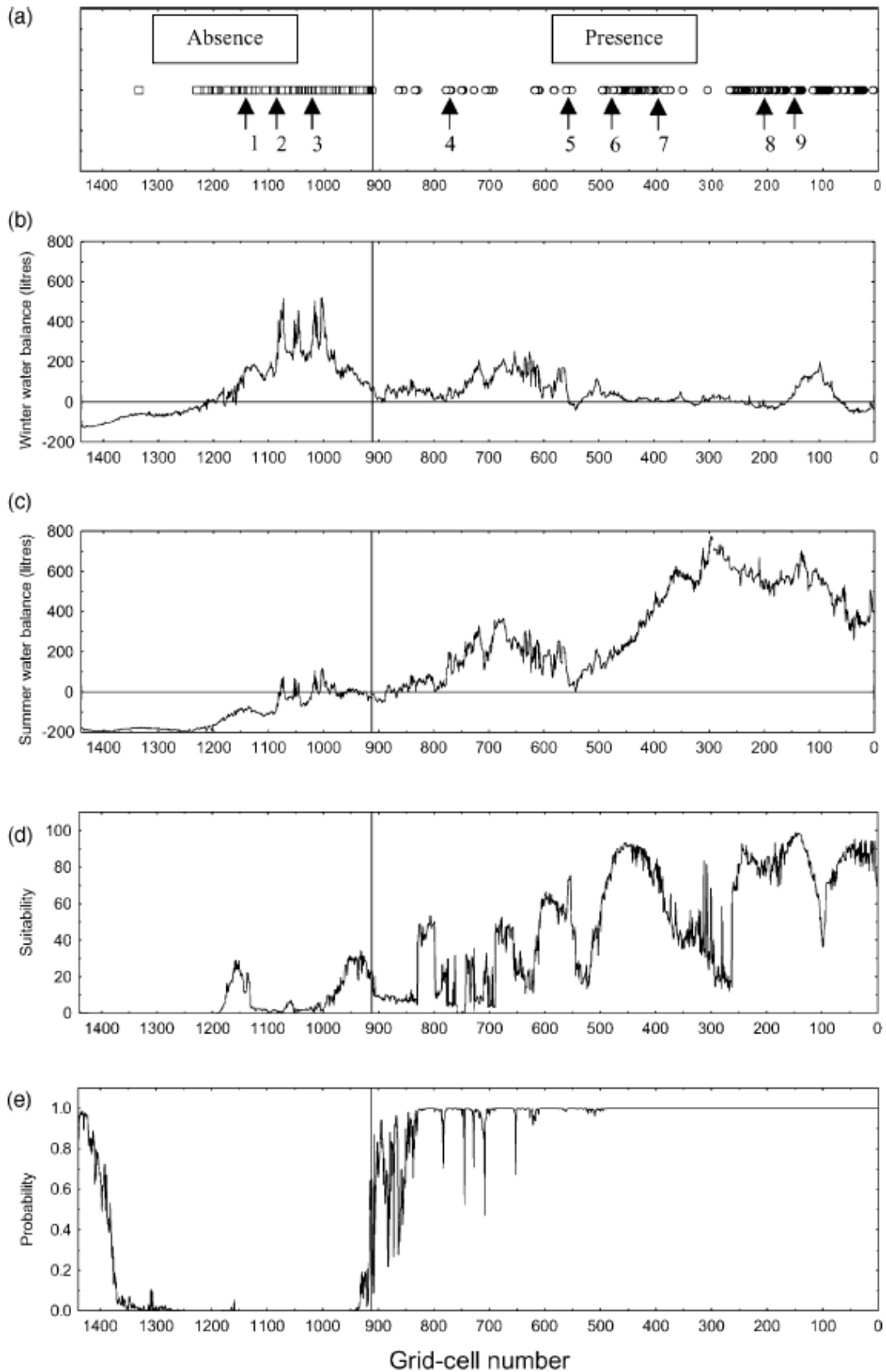
(a)

Absence | Presence

1 2 3 | 4 | 5 6 7 | 8 9

(b) Winter water balance (litres)

(c) Summer water balance (litres)

(d) Suitability

(e) Probability

Grid-cell number

Fig. 2. (a) The observed presence (*n*=211) and absence (*n*=76) of *S. plumieri* at localities along the South African coast from the Mozambique border (grid-cell 0) to the Namibian border (grid-cell 1439). The position of major towns and geographical features along the coast: 1: Cape Columbine; 2: Cape Town; 3: False Bay; 4: George; 5: Port Elizabeth; 6: Port Alfred; 7: East London; 8: Durban; 9: Mtunzini. The vertical line in all the figures indicates the position of Arniston, the most westward locality at which *S. plumieri* was observed present. The results of four predictive models: (b) winter water balance; (c) summer water balance; (d) suitability calculated using the PCA model; (e) probability calculated using the LR model.

Winter water balances (Fig. 2b) had low positive values to slight negative values on much of the east coast. Water balance values were particularly low between Port Alfred and Durban. On the south west coast between False Bay and Cape Columbine, relatively large water surpluses were evident. Localities on the west coast experienced relatively large water deficits.

In summer the south and east coasts were characterised by large water surpluses (Fig. 2c). Localities immediately to the east of Arniston (at the distribution limit) had small water surpluses or experienced small deficits. The west and south west coasts were characterised by large water deficits, particularly to the beyond Cape Town.

On the south and east coasts, predicted suitability (from the PCA model) was generally high but variable (Fig. 2d). Suitability on the west and south west coasts was lower than that on the south and east coasts with the exception of two peaks in suitability occurring near Cape Columbine and at localities just west of Arniston. The trend of lower suitability on the west and south west coasts and higher suitability on the south and east coasts follows the trend observed for SWB (Fig. 2c). A trough of low predicted suitability just to the east of Mtunzini (in the region of grid-cell 100) corresponds with a peak in winter water surplus (Fig. 2b).

The results of the LR model have been reported as probabilities which can be interpreted as probability of occurrence (Fig. 2e). The east coast was characterised by consistently high probabilities. The south coast was characterised by greater variability in the probability values with more localities having low probabilities. On the south west coast probabilities were consistently very low or zero. On the west coast probabilities were mostly zero but increased towards the Namibian border (grid-cell 1440). The coefficients, their associated standard errors and Wald statistics are presented in Table 4 for the LR model. Only linear terms were included in the linear predictor as scatter plots indicated that there was no justification for including higher order terms.

Table 4. Coefficients (*B*) with their associated standard errors (S.E.) and Wald statistics for the LR model

Variable	B	S.E.	Wald statistics
RN1	3.893	2.60E−08	2.25E+16
RN2	6.382	4.66E−08	1.88E+16
MNT1	−3.981	3.09E−08	1.66E+16
MXT1	3.723	3.53E−08	1.11E+16
RH1	−2.547	4.23E−08	3.63E+15
MXT2	−1.597	4.73E−08	1.14E+15
RH2	1.688	5.08E−08	1.10E+15
MNT2	−1.631	4.94E−08	1.09E+15
Constant	7.517	2.62E−07	8.23E+14

The full names of the variables can be found in Table 2.

3.1. Kappa statistics

Kappa statistics can be used to objectively assess the level of agreement between observed and predicted data. Monserud and Leemans (1992) suggested the following ranges of agreement for the κ statistic: no agreement, <0.05; very poor, 0.05–0.20; poor, 0.20–0.40; fair, 0.40–0.55; good, 0.55–0.70; very good, 0.70–0.85; excellent, 0.85–0.99; and perfect, 0.99–1.00. Negative values indicate extremely poor agreement (Monserud and Leemans, 1992). We used these ranges to describe the levels of agreement reported here using two tests.

Our first test involved using the locality records reserved for testing (the testing localities). The WWB model had a negative κ-value of −0.150 (Table 5) indicating extremely poor agreement between observed and predicted values (Monserud and Leemans, 1992). In contrast, the SWB model had a high κ-value (0.852) indicating excellent agreement (Table 5). The LR model had a κ-value of 1.000, indicating perfect agreement. The PCA model had a lower value (0.721) than that of SWB model, although this falls into the category of very good agreement.

Table 5. Tests of model performance using κ statistics and confusion matrix parameters

Model	Threshold	κ	a	b	c	d	n
WWB	0	−0.150	40	17	13	2	72
SWB	0	0.852	52	3	1	16	72
PCA	0.30	0.721	44	0	9	19	72
LR	0.13	1.000	53	0	0	19	72

Kappa statistics were calculated using thresholds that yielded maximum values for each of the models, with the exception of summer (SWB) and winter water balances (WWB) which were calculated using thresholds of zero. Kappa statistics were calculated using only those grid-cells in which *S. plumieri* was observed to be either present or absent (surveyed grid-cells).

Our second test involved using all of the predicted values from the PCA and LR models and measuring their agreement (using κ) with the SWB model. In order to test model agreement in this way, a set of "observed presence" and "observed absence" localities

(grid-cells) had to be defined. Grid-cells in the SWB model with water surpluses (≥0) were taken to represent "observed presence" and grid-cells with water deficits (0) were taken to represent "observed absence" to assess the level of agreement between the SWB model and the PCA and LR models (Table 6). There was good agreement between the SWB and the PCA model (0.679), and very good agreement between the SWB and the LR model (0.786; Table 6).

Table 6. Tests of model agreement using κ statistics and confusion matrix parameters calculated for the PCA and LR models using all the grid-cells along the entire coast

Model	Threshold	κ_{max}	a	b	c	d	n
PCA — all grid-cells	0.04	0.679	891	147	52	349	1439
LR — all grid-cells	0.89	0.786	827	29	116	467	1439

Calculations were performed as follows: those grid-cells with summer water balance values greater or equal to zero were taken to represent "observed presence" while grid-cells with water balance values below zero were taken to represent "observed absence". Thresholds that yielded maximum kappa (κ_{max}) values were selected.

4. Discussion

4.1. Interpretation of model predictions

We suggest that the predictions produced by each of these models may offer different insights into the potential distribution and biology of the target organism. The probabilities in the map generated from the LR model are interpreted as the probability of occurrence for the target species (*S. plumieri*). In contrast the probabilities in the map generated from the PCA model are interpreted as environmental suitability values (Robertson et al., 2001). The PCA model has no explicit "knowledge" of the conditions that exist at localities where the target organism is absent and thus the probabilities generated cannot be interpreted as probability of occurrence for the target organism. We use the term "suitability" to distinguish these from the probabilities produced using a LR model (Robertson et al., 2001). Similarly, the term "suitability" is also used to describe the values produced by another profile modelling technique based on ENFA (Hirzel et al., 2001 and Hirzel et al., 2002). The values of the water balance models are biologically meaningful since they empirically integrate energy and moisture levels.

Guisan and Zimmermann (2000) distinguish models that predict the fundamental niche from those that predict the realised niche of the target organism. Correlative models such as those presented here (PCA and LR) use actual distribution records to make predictions and these therefore must be drawn from the realised niche of that organism (Malanson et al., 1992). Thus, although biotic interactions are not explicitly accounted for (Robertson et al., 2001), their influence will be implicit by ampling the realised niche, and the result has been considered to be a prediction of the realised niche (Austin and Smith, 1989; Malanson et al., 1992; Franklin, 1995 and Guisan and Zimmermann, 2000). Austin (2002) has recently suggested that statistical models (correlative models) may not represent the realised niche but rather an amalgam of realised niche and sink areas. Sink areas are those areas where population growth is below replacement and populations are

maintained by dispersal from source areas, where population growth is positive (Pulliam, 1988).

In contrast, mechanistic models that are based only on physiological constraints and that do not explicitly account for biotic interactions (such as the SWB) tend to predict the fundamental niche of the target organism (Austin and Smith, 1989 and Guisan and Zimmermann, 2000). These can be refined to model the realised niche by adding rules to account for biotic interactions (see Prentice et al., 1992).

In the case of *S. plumieri*, there is probably very little difference between its realised and fundamental niche because it has few predators or pathogens, and effectively no competitors. This may help to explain the close agreement between predictions made by the mechanistic and correlative models in this study.

Mechanistic models (that are fundamental niche models) are considered to be most promising at successfully predicting climatically induced changes in the distribution of plant species (Malanson et al., 1992; Stephenson, 1998 and Guisan and Zimmermann, 2000). These models will be more robust under changed climatic conditions than correlative models as certain correlations may cease to apply under changed conditions (Prentice et al., 1992). In particular, correlative models have no way of handling the effects of climate change on the other organisms involved in the biotic interactions underlying realised niches. One solution is to incorporate certain biotic interactions into predictive models. For example, Leathwick et al. (1996) and Leathwick and Austin (2001) incorporated the effects of competition from a dominant species into models used to predict the spatial distribution of density of other species.

4.2. Model performance

There is fairly good visual agreement (Fig. 2) between the observed data in the form of presence and absence testing locality records and the SWB, the PCA and LR models. This was confirmed by κ statistics calculated using testing locality records (Table 5). In contrast, the κ statistic calculated for the WWB model was negative (Table 5), which indicates extremely poor agreement between the model and the testing localities (Monserud and Leemans, 1992). Using the scale of agreement proposed by Monserud and Leemans (1992), the κ statistics indicated "very good" agreement for the PCA and "excellent" agreement for the SWB model and "perfect" agreement for the LR model.

Although the κ-values calculated for the LR model indicated "perfect" agreement with the testing locality records, this probably represents an overestimate of the actual performance of the LR model. The LR model predicted high probabilities of *S. plumieri* being present along the west coast close to the Namibian border (between grid-cells 1470 and 1440; Fig. 2e). However, we believe that *S. plumieri* is unlikely to occur in this region. Unfortunately, we were unable to evaluate the performance of the LR model quantitatively (using κ statistics) along this section of the west coast due to a lack of testing localities (Fig. 2a). Two sources of indirect evidence suggest that *S. plumieri* is unlikely to occur in this region. Firstly, the LR model suggests that *S. plumieri* is likely to

occur at the Namibian border (grid-cell 1440 at the edge of the map region), which in turn suggests that it is also likely to occur just beyond the current map region, along the arid coast of Namibia. However, no herbarium records exist for this species from Namibia, and a distribution map produced by Tinley (1985) suggests that *S. plumieri* does not occur in Namibia but only occurs considerably further north along the northern coast of Angola. Finally, the SWB and PCA models also suggest that *S. plumieri* is unlikely to occur along this section of the South African coast. Although we believe *S. plumieri* to be absent along this section of the coast, a survey is required to confirm this.

The LR model attained a higher κ-value than the SWB model, suggesting better agreement between the observed and predicted values for the LR model than for the SWB model. This may be at least partly influenced by the method of evaluation used. Chatfield (1995) maintains that splitting data into a training and a testing set is a poor substitute for true replication, as the two datasets are not completely independent. This is likely to be true in this case and this dependence (between the training and testing sets) could explain the higher κ-value calculated for the LR model than that calculated for the SWB model. The LR model was built using a training set of presence and absence localities, and then evaluated using a set of testing localities that were not completely independent of the training set. In contrast, the SWB model was based on empirical ecophysiological data rather than on presence and absence localities. This model was thus evaluated using data that were independent of the data used to build it. The LR model thus had a better chance of performing well using this model evaluation test than the SWB model. Similarly, the PCA model also had a better chance of performing well using this test than the SWB model—the only difference being that the PCA model was built using only presence data (rather than presence and absence data as was the case with the LR model). This is the reason for using the *model agreement* tests (Table 6).

Kappa-values suggest that the LR model also performed better than the PCA model. This is possibly because the LR model had the advantage of being built using 57 absence localities in addition to the 158 presence localities used in the PCA model.

The performance of the mechanistic model would probably have been better had another mechanistic process been incorporated into this model. Although the LR and PCA models had a greater chance of performing well using these model evaluation tests than the SWB model, all three models demonstrated good performance.

4.3. Model agreement

Kappa-values that were calculated using the reclassified SWB map were used to assess the level of agreement between the SWB and both the PCA and LR models. Kappa-values indicated "good" agreement between the SWB model and the PCA model (Table 6). The agreement between the SWB model and the LR model was "very good" using the scale of agreement proposed by Monserud and Leemans (1992). These values suggest very good correspondence between predictions made using a simple mechanistic model (the SWB model) and two correlative models (the PCA and LR models). Slightly better agreement between the LR model and the SWB model than between this model and the

PCA model can again be explained by the LR model having 57 absence localities in addition to the 158 presence localities used to build the PCA model. The close agreement between the correlative (PCA and LR) models and the mechanistic model (SWB) may be because the realised niche of *S. plumieri* probably quite closely resembles its fundamental niche.

4.4. Water balance as a predictor variable in correlative models

The results of model performance tests suggest that the SWB model is a far better predictor of *S. plumieri* presence than the WWB model. This may be explained by considering the phenology of the plant (Peter et al., 2003). The SWB model was calculated for those months which coincide with periods when the plant is actively growing and reproducing (Peter et al., 2003). This suggests that water balance values calculated for these periods may be mechanistically more important, and as a result should be more important and useful for predicting plant distributions. This is particularly important when water balance is used as a predictor variable in correlative models (in contrast to the mechanistic approach adopted in this study). Water balance has been used as a predictor variable in various correlative studies (for example, Leathwick, 1995; Leathwick, 1998; Leathwick et al., 1996 and Leathwick and Whitehead, 2001). Stephenson (1998) suggested that site water balance should be used as a predictor variable for the purposes of building correlative models, even if it is only crudely calculated. Site water balance considers the interactions of energy and water which are important for predicting distributions (Stephenson, 1998). In addition, Guisan and Zimmermann (2000) have suggested that "physiology-based" parameters such as site water balance should be used in preference to physiographic predictors for developing static models that are more mechanistic. Our data suggest that the phenology of the target species should be taken into consideration when calculating water balance values in order to focus on those periods which are biologically significant.

4.4.1. Profile versus group discrimination techniques

An issue that is likely to generate considerable debate in the future is the choice between profile and group discrimination techniques for predicting species distributions. Profile models are likely to be appropriate when absence data are not available or are unreliable. In a recent study, using simulated data, Hirzel et al. (2001) found that a profile technique (ENFA) performed better than a group discrimination technique (GLM) in cases where the target organism was not in equilibrium with the environment, e.g. alien invasive organisms. However, when reliable presence and absence data are available then group discrimination techniques are likely to perform best (Ferrier and Watson, 1997 and Hirzel et al., 2001). This is confirmed by the slightly better agreement between the SWB and LR models than the agreement between the SWB and PCA models.

When a target organism is not in equilibrium with its environment, absence data collected for this organism are likely to contain a proportion of false absence records, which may adversely affect model performance (Hirzel et al., 2001). However, logistic regression has been successfully applied using "unsurveyed absence" records (e.g. Cumming, 2000a

and Cumming, 2000b). Recently, Zaniewski et al. (2002) successfully used "pseudo-absence" data to apply GAM for predicting the distribution of ferns.

There appears to be a need for further comparative studies that compare the performance of profile and group discrimination techniques using data of varying quality. In particular, an important question is whether group discrimination techniques such as GLM or GAM can be applied without using surveyed absence data (e.g. Cumming, 2000a and Cumming, 2000b; Zaniewski et al., 2002) to produce models that perform better than those produced by profile techniques (e.g. Robertson et al., 2001 and Hirzel et al., 2002), given data of the same quality.

4.5. Conclusion

The results suggest that correlative models can perform as well or better than simple mechanistic models. However, the generality of this statement requires testing. These models have different requirements in terms of input data and prior knowledge of the target organism's biology. The predictions generated from these models are each likely to offer slightly different insights into the potential distribution and biology of the target organism and may be appropriate for different purposes. Mechanistic models such as those used here (the SWB model) require a greater knowledge of the biology of the organism and require making time-consuming ecophysiological measurements. These models may in turn yield greater insights into the biology of the organism that are mechanistic in nature. In contrast, correlative models are easier and less time-consuming to build and require less prior knowledge of the target organism's biology. These models will yield different insights into the biology and potential distribution of the target organism from mechanistic models. The choice of model is likely to be influenced by several factors, such as the aims of the study, the biology of the target organism, the level of knowledge the target organism's biology and data quality. We suggest that the type of target organism and the level of knowledge of the target organism's biology will play a central role in the model development process. Simple correlative models may be used initially when the biology and distribution of the organism is not well known. These models can then be iteratively refined (see Chatfield, 1995) to produce models that are more mechanistic in nature. This is equivalent to incorporating more knowledge of "ecological process" (Austin, 2002) into the model. This may be followed (if feasible) by the development of ecophysiological models that simulate the mechanisms considered to underlie the correlations between the distribution records and the predictor variables observed in developing the correlative models.

Acknowledgements

References

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* **157**, pp. 101–118.

Austin, M.P. and Meyers, J.A., 1996. Current approaches to modelling the environmental niche of eucalypts: implications for management of forest biodiversity. *For. Ecol. Manage.* **85**, pp. 95–106.

Austin, M.P. and Smith, T.M., 1989. A new model for the continuum concept. *Vegetatio* **83**, pp. 35–47.

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

Austin, M.P., Nicholls, A.O. and Margules, C.R., 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Monogr.* **60**, pp. 161–177.

Austin, M.P., Nicholls, A.O., Doherty, M.D. and Meyers, J.A., 1994. Determining species response functions to an environmental gradient by means of a beta-function. *J. Veg. Sci.* **5**, pp. 215–228.

Beerling, D.J., Huntley, B. and Bailey, J.P., 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *J. Veg. Sci.* **6**, pp. 269–282

Buckland, S.T. and Elston, D.A., 1993. Empirical models for the spatial distribution of wildlife. *J. Appl. Ecol.* **30**, pp. 478–495.

Busby, J.R., 1991. BIOCLIM—a bioclimatic analysis and prediction tool. In: Margules, C.R., Austin, M.P. (Eds.), Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Melbourne, pp. 64–68.

Caithness, N., 1995. Pattern, process and the evolution of the African antelope (Mammalia: Bovidae). Ph.D. Dissertation, University of the Witwatersrand, Johannesburg, 210 pp.

Carpenter, G., Gillison, A.N. and Winter, J., 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiv. Conserv.* **2**, pp. 667–680.

Chatfield, C., 1995. Model uncertainty, data mining and statistical inference. *J. R. Statist. Soc. A* **158**, pp. 419–446

Cumming, G.S., 2000. Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida). *J. Biogeogr.* **27**, pp. 425–440.

Cumming, G.S., 2000. Using between-model comparisons to fine-tune linear models of species ranges. *J. Biogeogr.* **27**, pp. 441–455.

Erasmus, B.F.N., Kshatriya, M., Mansell, M.W., Chown, S.L. and Van Jaarsveld, A.S., 2000. A modelling approach to antlion (Neuroptera: Myreleontidae) distribution patterns. *Afr. Entomol.* **8**, pp. 157–168.

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, p. 193.

Fielding, A.H. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, pp. 38–49.

Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* **19**, pp. 474–499.

Guisan, A. and Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, pp. 147–186.

Guisan, A., Theurillat, J.-P. and Kienast, F., 1998. Predicting the potential distribution of plant species in an alpine environment. *J. Veg. Sci.* **9**, pp. 65–74.

Guisan, A., Weiss, S.B. and Weiss, A.D., 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecol.* **143**, pp. 107–122.

Higgins, S.I., Richardson, D.M. and Cowling, R.M., 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conserv. Biol.* **13**, pp. 303–313.

Hirzel, A.H., Helfer, V. and Métral, F., 2001. Assessing habitat-suitability models with a virtual species. *Ecol. Model.* **145**, pp. 111–121.

Hirzel, A.H., Hausser, J., Chessel, D. and Perrin, N., 2002. Ecological-Niche Factor Analysis: how to compute habitat-suitability maps without absence data?. *Ecology* **83**, pp. 2027–2036.

Huberty, C.J., 1994. Applied Discriminant Analysis. Wiley/Interscience, New York, 466 pp.

Jackson, D.A., 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**, pp. 2204–2214.

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, p. 99.

Leathwick, J.R., 1995. Climatic relationships of some New Zealand forest tree species. *J. Veg. Sci.* **6**, pp. 237–248.

Leathwick, J.R., 1998. Are New Zealand's *Nothofagus* species in equilibrium with their environment?. *J. Veg. Sci.* **9**, pp. 719–732.

Leathwick, J.R. and Austin, M.P., 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* **82**, pp. 2560–2573.

Leathwick, J.R. and Mitchell, N.D., 1992. Forest pattern, climate and vulcanism in central North Island, New Zealand. *J. Veg. Sci.* **3**, pp. 603–616.

Leathwick, J.R. and Whitehead, D., 2001. Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Funct. Ecol.* **15**, pp. 233–242.

Leathwick, J.R., Whitehead, D. and McLeod, M., 1996. Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modelling approach. *Environ. Software* **11**, pp. 81–90.

Lees, B.G., 1994. Decision trees, artificial neural networks and genetic algorithms for classification of remotely sensed and ancillary data. *7th Aust. Remote Sens. Conf. Proc.* **1**, pp. 51–59.

Malanson, G.P., Westman, W.E. and Yan, Y.-L., 1992. Realized versus fundamental niche functions in a model of chaparral response to climatic change. *Ecol. Model.* **64**, pp. 261–277

Manel, S., Dias, J.-M. and Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: as case study with a Himalayan river bird. *Ecol. Model.* **120**, pp. 337–347

McCullagh, P., Nelder, J.A., 1983. Generalized Linear Models. Chapman and Hall, London, p. 261.

Michaelsen, J., Schimel, D.S., Friedl, M.A., Davis, F.W. and Dubayah, R.C., 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *J. Veg. Sci.* **5**, pp. 673–686.

Monserud, R.A. and Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. *Ecol. Model.* **62**, pp. 275–293.

Neilson, R.P., 1995. A model for predicting continental-scale vegetation distribution and water balance. *Ecol. Appl.* **5**, pp. 362–385.

Nicholls, A.O., 1989. How to make biological surveys go further with generalised linear models. *Biol. Conserv.* **50**, pp. 51–75.

Nix, H.A., 1986. A biogeographical analysis of Australian elapid snakes. In: Longmore, R. (Ed.), Atlas of Elapid Snakes of Australia. Australian Government Publishing Service, Canberra, pp. 4–15.

Osborne, P.E. and Tigar, B.J., 1992. Interpreting bird atlas data using logistic models: an example from Lesotho, southern Africa. *J. Appl. Ecol.* **29**, pp. 55–62.

Packer, M.J., Canney, S.M., McWilliam, N.C., Abdallah, R. 1999. Ecological mapping of a semi-arid savanna. In: Coe, M.J., McWilliam, N.C., Stone, G.N., Packer, M.J. (Eds.), Mkomazi: The Ecology, Biodiversity and Conservation of a Tanzanian Savanna. Royal Geographical Society (with the Institute of British Geographers), London, pp. 43–68.

Palmer, A.R. and Van Staden, J.M., 1992. Predicting the distribution of plant communities using annual rainfall and elevation: an example from southern Africa. *J. Veg. Sci.* **3**, pp. 261–266.

Pearce, J. and Ferrier, S., 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecol. Model.* **128**, pp. 127–147.

Peter, C.I. and Ripley, B.S., 2000. An empirical formula for estimating the water use of *Scaevola plumieri*. *S. Afr. J. Sci.* **96**, pp. 1–4.

Peter, C.I., Ripley, B.S. and Robertson, M.P., 2003. Environmental limits to the distribution of *Scaevola plumieri* along the South African coast. *J. Veg. Sci.* **14**, pp. 89–98.

Prentice, I.C., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. and Solomon, A.M., 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* **19**, pp. 117–134.

Pulliam, H.R., 1988. Sources, sinks, and population regulation. *Am. Nat.* **132**, pp. 652–661.

Robertson, M.P., Caithness, N. and Villet, M.H., 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Div. Distrib.* **7**, pp. 15–27.

Robinson, T.P., Rogers, D.J. and Williams, B.G., 1997. Mapping tsetse habitat suitability in the common fly belt of southern Africa using multivariate analysis of climate and remotely sensed vegetation data. *Med. Vet. Entomol.* **11**, pp. 235–245.

Rogers, D.J. and Randolph, S.E., 1993. Distribution of tsetse and ticks in Africa: past, present and future. *Parasitol. Today* **9**, pp. 266–271.

Rogers, D.J., Williams, B.G., 1993. Tsetse distribution in Africa: seeing the wood and the trees. In: Edwards, P.J., May, R. (Eds.), Large-Scale Ecology and Conservation Biology. Blackwell Scientific Publications, Oxford, pp. 247–271.

Rogers, D.J., Hay, S.I. and Packer, M.J., 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Ann. Trop. Med. Parasitol.* **90**, pp. 225–241.

Schulze, R.E., Maharaj, M., Lynch, S.D., Howe, B.J., Melvil-Thomson, B., 1997. South African Atlas of Agrohydrology and Climatology, 1st ed. Water Research Commission, Pretoria.

Stephenson, N.L., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *J. Biogeogr.* **25**, pp. 855–870.

Tinley, K.L., 1985. Coastal dunes of South Africa. South African National Scientific Programmes, Report no. 109.

Walker, P.A., 1990. Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate. *J. Biogeogr.* **17**, pp. 279–289.

Williams, B.G., Rogers, D.J., Staton, G., Ripley, B., Booth, T., 1994. Statistical modelling of georeferenced data: mapping tsetse distributions in Zimbabwe using climate and vegetation data. In: Perry, B.D., Hansen, J.W. (Eds.), Modelling Vector-Borne and Other Parasitic Diseases. ILRAD, Nairobi, pp. 267–280.

Woodward, F.K. and Williams, B.G., 1987. Climate and plant distribution at global and local scales. *Vegetatio* **69**, pp. 189–197.

Zaniewski, A.E., Lehmann, A. and Overton, J.McC., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecol. Model.* **157**, pp. 261–280.