

Developing an approach to defining the potential distributions of invasive plant species: a case study of *Hakea* species in South Africa

David C. Le Maitre^{1*}, Wilfried Thuiller² and Lucille Schonegevel¹

¹Natural Resources and Environment CSIR, PO Box 320, Stellenbosch, South Africa, ²Laboratoire d'Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France

ABSTRACT

Aim Models of the potential distributions of invading species have to deal with a number of issues. The key one is the high likelihood that the absence of an invading species in an area is a false absence because it may not have invaded that area yet, or that it may not have been detected. This paper develops an approach for screening pseudo-absences in a way that is logical and defensible.

Innovation The step-wise approach involves: (1) screening environmental variables to identify those most likely to indicate conditions where the species cannot invade; (2) identifying and selecting the most likely limiting variables; (3) using these to define the limits of its invasion potential; and (4) selecting points outside these limits as true absence records for input into species distribution models.

This approach was adopted and used for the study of three prominent *Hakea* species in South Africa. Models with and without the false absence records were compared. Two rainfall variables and the mean minimum temperature of the coldest month were the strongest predictors of potential distributions. Models which excluded false absences predicted that more of the potential distribution would have a high invasion potential than those which included them.

Main conclusions The approach of applying a priori knowledge can be useful in refining the potential distribution of a species by excluding pseudo-absence records which are likely to be due to the species not having invaded an area yet or being undetected. The differences between the potential distributions predicted by the different models convey more information than making a single prediction, albeit a consensus model. The robustness of this approach depends strongly on an adequate knowledge of the ecology, invasion history and current distribution of that species.

Keywords

False absence, *Hakea*, limiting factors, potential distribution, presence, Proteaceae, pseudo-absence, South Africa, species distribution models, true absence.

*Correspondence: David Le Maitre, Natural Resources and Environment, CSIR, PO Box 320, Stellenbosch, South Africa. E-mail: dLmaitre@csir.co.za

INTRODUCTION

The modelling of species distributions using statistically based models is a growing area of research. There are strongly divergent schools of thought on both the approaches and methods, and about how to deal with the assumptions and uncertainties in interpreting the outputs. Applications of these models to predict plant species distributions have to consider and deal with several critical issues and uncertainties which include: presence-only data, the assumption that the boundaries are the outcome of a dynamic equilibrium between the species and the environment, the assumption that a distribution is determined primarily by abiotic factors, and data structure issues such as spatial autocorrelation in the variables and varying spatial resolution of different data sets (Guisan & Zimmermann, 2000; Robertson *et al.*, 2001; Guisan & Thuiller, 2005; Thuiller *et al.*, 2005; Barry & Elith, 2006; Guisan *et al.*, 2006; Pearson *et al.*, 2006). Some studies have also recognized the problems associated with the inclusion of false absences (Hirzel *et al.*, 2001; Lütolf *et al.*, 2006). Presenceonly modelling approaches offer an alternative approach to this problem (Hirzel *et al.*, 2002) but may overestimate potential distributions compared with presence–absence models (Zaniewski *et al.*, 2000; Brotons *et al.*, 2004) although the range of models and their performance is improving (Elith *et al.*, *al.*, 2006). This study focuses on developing a methodology for producing more robust models by selectively excluding false absences. The approach entails screening environmental factors to identify those that appear most likely to be determining the distributions of the species of interest. Thresholds in these factors are then used to divide pseudo-absence records into those that are likely to be false absences and those likely to be true absences so that only the true absences are used in modelling.

There are three principal reasons why the modelling of the potential distributions of invading species is problematic. The first problem is common to most studies of species distributions and is caused by the fact that the data confirm presence but not absence (Green, 1971). There are three main reasons why a species may be absent from a locality (Green, 1971; Pulliam, 2000; Hirzel et al., 2001, 2002; Lütolf et al., 2006): (1) it cannot establish a persistent population because abiotic or biotic factors prevent successful reproduction; (2) it could potentially establish a population but has not been able to reach that locality yet (e.g. dispersal limitation); or (3) it is present but was not detected or recorded, possibly because it is only present as seeds. The first case would be a true absence and would lie outside the species' potential niche. The probability of (2) or (3) contributing to false absences is likely to be higher for an invading than a native species because the invader may not be easily recognized as new, is not likely to have been resident for long and may be sparse and difficult to detect.

The second problem is the assumption that the species is in quasi-equilibrium with its environment, i.e. the boundaries of its current distribution are those of its realized niche sensu Hutchinson (1957) (Green, 1971; Austin, 2002). This assumption is unlikely to be correct for many invasive species (Hirzel et al., 2001, 2002; Peterson, 2003; Guisan & Thuiller, 2005; Hierro et al., 2005), particularly given that: (1) most invasive species in South Africa were introduced within the last 100-150 years (e.g. Shaughnessy, 1986); (2) that those introductions were generally confined to just a few localities; and (3) that even species with long-range dispersal have few propagules dispersing more than a kilometre (Nathan et al., 2002; Higgins et al., 2003; Midgley et al., 2006). Given these limitations, it is highly unlikely that the species has achieved its realized niche in every area it has invaded; in some areas it may have reached its limits, in others it may not (Welk, 2004; Mau-Crimmins et al., 2006). Nevertheless, and contrary to Hirzel et al. (2002), we believe that modelling the climatic envelope of an invading species is useful because data on areas that are likely to be highly suitable are more useful than data on where it is has a low probability of invading (Thuiller et al., 2005; Richardson & Thuiller, 2007). Although initial introductions and the processes from naturalization to invasion do have an element of chance, the subsequent invasion processes tend to generate regular patterns that can be used in analysing species distributions (Arim et al., 2006; Thuiller et al., 2006).

The third issue is the assumption that abiotic factors are the primary determinants of a species distribution and that biotic factors – including competition, pathogens and facilitatory relationships – are relatively unimportant (Keane & Crawley, 2002; White *et al.*, 2006). However, a number of species have

demonstrated an ability to invade novel environments that do not closely match those of the environment they evolved in, examples being a number of pine species (Richardson & Bond, 1991). There are two main hypotheses about why this occurs: (1) small populations of the species that are introduced are able to evolve rapidly to become invasive (Lavergne & Molofsky, 2007) due to, for example, rapid adaptation and selection to overcome Allee effects (Taylor & Hastings, 2005); or (2) the primary factor is either (i) the species is inherently a superior competitor or (ii) it is the release from inhibiting factors (biotic inhibitors) such as pathogens, pests, better competitors and predators which gives the species a competitive advantage (Keane & Crawley, 2002; Vilà & Weiner, 2004; Hierro et al., 2005; Callaway & Maron, 2006). However, some studies have not found a significant degree of release, especially from generalist herbivores (Joshi & Vrieling, 2005; Parker et al., 2006). Although rapid evolution may be important, especially in species with marked lags between introduction and invasion (Pyšek & Hulme, 2005), we suggest that for most species the release from inhibiting biotic factors is the most important factor, and may often be the cause of their apparently superior competitive ability. This rationale is supported by the successes that have been achieved in biocontrol where the introduction of carefully selected pests or pathogens has often turned a major invader into a minor one or even reversed its invasions (Hoffmann & Moran, 1998; Keane & Crawley, 2002). If biotic inhibitors generally are the primary determinants of species distributions, this suggests that limits on the distribution of relatively pest- and pathogen-free invading species should be determined primarily by their climatic tolerances, i.e. they more closely approach the fundamental niche (sensu Hutchinson, 1957) of the species, although interspecific competition may still be important.

At least five Hakea species have been introduced to South Africa and four species have become important invaders (Table 1). This paper only focuses on the major invaders in the fynbos biome: Hakea sericea, Hakea gibbosa and Hakea drupacea. Hakea drupacea and H. gibbosa have invaded relatively limited areas compared with H. sericea (Fig. 1). The reasons for these differences are not entirely clear. All three species were introduced more than 150 years ago with H. sericea being the last to be introduced (Table 1). Active promotion of H. drupacea for hedges probably began in the 1850s and was maintained until at least the early 1900s (Shaughnessy, 1986). The most widely distributed species, H. sericea, was introduced last, was promoted the least and had few seeds sold (Kruger et al., 1986; Shaughnessy, 1986), yet it has an order of magnitude more populations and the greatest distribution range (Table 2) with nearly 40% of all populations east of 20° E. These patterns support the hypothesis that although people played a key role in introducing H. sericea, its subsequent spread was largely due to its inherent invasiveness (Fugler, 1979; Richardson, 1984; Kruger et al., 1986; Shaughnessy, 1986), primarily due to its short juvenile period and high seed production (Table 1; Richardson et al., 1987).

A number of studies have suggested that home (native) range data can be useful when predicting potential distributions (e.g.

Table 1 Ecological information on four of the Hakea species introduced to South Africa. Data collated from Neser (1983a,b); Neser & Fugl	r
(1983); Shaughnessy (1986); Richardson et al. (1987) and the South African Plant Invaders Atlas (SAPIA) data base (Henderson, 1998).	

Species	Date of Reason for, and introduction extent of, dissemination		Vegetation type invaded (biome)	SAPIA records & percentage of 0.25° squares in fynbos	Seed bank size, seed survival of fire, juvenile period	
Hakea sericea Schrad.	1858	Hedges, accidental; not widely planted	Fynbos: montane, grassy Grassland: coastal grassland*	228 W and E Cape 1 KwaZulu-Natal 64%	Large; moderate; 2–3 years	
Hakea drupacea (Gaertn. F.) Roemer & Schultes (H. suaveolens R.Br.)	1850	Hedges; widely planted	Fynbos: montane, grassy, lowland	58 W and E Cape 17%	Small; moderate; 6 years	
Hakea gibbosa (Sm.) Cav.	1835	Hedges, sand stabilization; moderately widely planted	Fynbos: montane, grassy	34 W and E Cape 11%	Small; high; 2 years	
Hakea salicifolia (Vent.) Burtt.	1835?	Hedges; widely planted	Fynbos Grassland	3 W Cape 2 KwaZulu-Natal 1.6%	Intermediate; low; not given	

*The Pondoland coastal grassland occurs on a geological formation which is part of the same group that is dominant in the fynbos biome and the vegetation includes many species with phylogenetic relationships with fynbos (van Wyk & Smith, 2001).

Table 2 Summary of the distributions of three Hakea species based on quarter-degree square data (0.25° latitude × longitude, QDS) from theSouth African Plant Invaders Atlas data base (SAPIA; Henderson, 1998) and records from the Protea Atlas data base (A. G. Rebelo, personalcommunication, 2006).

Species	SAPIA records number of QDSs	Protea Atlas records number of 1' squares	Localities (QDS) recorded in SAPIA but not in Protea Atlas	Localities (QDS) recorded in Protea Atlas but not in SAPIA
Hakea drupacea	58 29	518 171	4	1
Hakea gibbosa	36 21	765 221	3	1
Hakea sericea	226 85	6005 1719	5	18
	Easternmost (SAPIA*)	Easternmost (Protea Atlas)	Mean location (Protea Atlas data)	Range (deg latitude)
Hakea drupacea	26.12° E, 33.38° S	25.58° E, 32.69° S	18.89° E, 34.20° S	7.26
Hakea gibbosa	26.63° E, 33.38° S	25.55° E, 32.84° S	19.27° E, 34.31° S	7.21
Hakea sericea	30.12° E, 30.88° S	30.19° E, 30.93° S	20.58° E, 33.88° S	11.86

*Calculated as the centre of the quarter-degree square.

Welk, 2004; Hierro *et al.*, 2005) but others have found that they add little value (Mgidi *et al.*, 2007) although they may provide useful insights (Mau-Crimmins *et al.*, 2006). There are descriptions of the native distribution of the *Hakea* species in Australia (Flora of Australia Online: http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/main/, accessed 13 November 2006) but they do not include actual locality records and are not suitable for this study.

A key characteristic of all three invasive *Hakea* species is the large crop of seeds which is retained in the serotinous fruit until a fire kills the parent plant, releasing the seeds for dispersal by the wind (Richardson *et al.*, 1987). The results of a seed-dispersal simulation model for *H. sericea* (Le Maitre *et al.* 2008) suggest that many seeds can disperse more than 100 m and sufficient seeds disperse 500 m or more from the parent for very rapid range expansions after fires. The evidence of rapid population

growth and spread by *H. sericea* suggests strongly that its current distribution in mountain areas in the vicinity of Cape Town, where it has been present for more than 100 years, is not limited by seed dispersal and therefore approximates the environmental envelope. However, the distributions of *H. gibbosa* and *H. drupacea* are more restricted and may not approximate their potential environmental envelope.

INNOVATION

Developing an approach for filtering pseudo-absences

The problem with generating spatially random pseudo-absence data for use in presence–absence modelling techniques is that a given sample of random spatial pseudo-absence points is likely to include some points which fall within the range of the climate



Figure 1 Locality map and the current distributions of three Hakea species in South Africa based on records from the Protea Atlas data base (A. G. Rebelo, personal communication, 2006).

values found in the presence data. Records from inside the limits of climate values found in the presence data are potentially false absences, at least in terms of the climatic variables being assessed and should, ideally, be excluded from the pseudo-absence data. This rationale is followed in this paper where we compare models developed using data sets which include or exclude absence samples from within ranges of climatic factors believed to represent true limits of the potential distribution of the invading species. The first step is to screen potential variables using general ecological principles, and knowledge of the species, to identify those that are most likely to be limiting. The second step is to use thresholds in one of more of these variables to divide the areas where there are no presence records into: (1) those that are probably invadable because they fall within those thresholds (false absences); and (2) those considered noninvadable because they fall outside those thresholds (true absences). The variables used to define the thresholds must be excluded from the subsequent modelling because stepwise models would select them because they are directly linked to the presence and absence thresholds.

Study area

The potential study area was the whole area covered by the available climate data sets, namely South Africa, Lesotho and Swaziland (Fig. 1). The selection of threshold values for the climatic variables reduced this total area substantially depending on the particular variables. A $1' \times 1'$ latitude × longitude grid (based on the WGS 1984 datum) was used as the basic grid for the analyses and modelling in this study and is referred to as the 1MS grid.

Species distribution data

Data on the spatial distribution of Hakea species in South Africa are available from the South African Plant Invaders Atlas (SAPIA) (Henderson, 1998) as well as the Protea Atlas (Rebelo, 1991). There are important differences between these data sets in species distributions and numbers of records (Table 2). The Protea Atlas data were selected because they are available at a much finer resolution (< 1' of latitude \times longitude) and generally give a more thorough coverage (Table 2). More than 7400 records are available for the three Hakea species used in this study: 518 for H. drupacea, 765 for H. gibbosa and more than 6000 for H. sericea. Population sizes at each locality were recorded in four classes: 1-9 individuals, 10-100, 100-10,000 and > 10,000 plants. Populations could also be recorded as extinct where there were dead adults but no regeneration. 'Extinct' records were included in this study because they indicate that the species had been able to invade this site. The value of each climatic variable for each Hakea atlas record was obtained by using the Identity function in ArcGIS to overlay the presence (point) data on the underlying gridded climatic data.

Two presence data sets were used in the analyses. The complete presence data were used in the determination of thresholds and initial analyses. For the regression model fitting, a random sample of 10% of the presence records for each species was selected using the random sampling option in the Animal Movement extension developed for ArcView 3.2 (Hooge & Eichenlaub, 1997; http://www.absc.usgs.gov/glba/gistools/animal_mvmt.htm; copy of Version 2.0 downloaded 2 March 2007). This was done to reduce the number of true absence samples (see below) required to provide similar numbers of presence and absence records, i.e. moderate prevalence as recommended by McPherson *et al.* (2004), whilst balancing this with the general finding that increasing sample sizes result in asymptotic improvements in model accuracy (Stockwell & Peterson, 2002; Kadmon *et al.*, 2003; Reese *et al.*, 2005; Heikkinen *et al.*, 2006; but see Elith *et al.*, 2006). In this case there were at least 100 presence and absence records for each analysis.

Absence data

A set of 1500 1MS grid cells from the entire study area was selected using random coordinates. A random subsample of 500 of these was chosen for use in the models which included all absences (i.e. true and false absences). A second random set of 500 absence records was selected after the original 1500 had been filtered to exclude those with values that lay within the predetermined thresholds. These were used as the true absences.

Statistical analyses and modelling

Statistical analyses and the generation of the sample of 1500 coordinates, for selection of true absence locations, were done using SAS version 9.1.2 (SAS Institute Inc., 2004) and Microsoft® Office Excel 2003. The processing of the spatial data was done using ArcView 3.2a and ArcGIS 9.1. The distribution model fitting was done with the BIOMOD package (version 2006.01.26) developed by one of the authors (Thuiller, 2003) and implemented within the R software (http://cran.r-project.org/; copy of Version 2.3.1 (2006-06-01) downloaded 20 August 2006). The BIOMOD package fits a range of statistical models to species presences and absences and the values for environmental variables for the same localities. The model mainly used in this analysis was the generalized linear model (GLM). The results of the generalized additive model (GAM) were generally very similar. BIOMOD also fits a surface range envelope (SRE) which simply uses the upper and lower limits of each of the climate variables in the presence records to define a multidimensional envelope. Potentially nonlinear responses to climatic variables were included by allowing both GLM and GAM models to fit low-order polynomial relationships.

The goodness of fit was evaluated using the residual deviance and the Akaike information criterion, the predictive accuracy was estimated using the area under the receiver operating curve (AUC) and the presence–absence cut-off was selected as the one maximizing the percentage of presences and true absences which were correctly predicted. BIOMOD randomly partitioned the data (70% model calibration and 30% for model evaluation) as recommended by Guisan & Zimmermann (2000). **Table 3** Summary of the chi-square tests for the basic rainfall and temperature variables from Schulze *et al.* (1997), Mgidi *et al.* (2007) and from variables developed by Schulze & Perks (1999) and used by Rutherford *et al.* (2000). The number of classes used was based on the range of the data and an analysis of the distribution of all the data relative to the cells where *Hakea* occurred. Equal size classes were used in most cases. All chi-square values are significant (P > 0.05).

Variable	Chi-square value
Mean annual rainfall (mm)	2790.28
Rainfall of wettest month (mm)	1796.47
Rainfall of driest month (mm)	5365.57
Rainfall of wettest quarter (mm)	2530.36
Rainfall of driest quarter (mm)	7624.97
Rainfall of coolest quarter (mm)	10854.85
Rainfall of warmest quarter (mm)	1145.89
Rainfall concentration (%)	1966.18
Mean annual temperature (°C)	516.66
Average temperature range (°C)	1763.04
Minimum temperature of coolest month (°C)	1854.92
Maximum temperature of warmest month (°C)	651.43
Mean temperature of coolest quarter (°C)	759.31
Mean temperature of warmest quarter (°C)	843.25
Mean temperature of wettest quarter (°C)	2495.38
Mean temperature of driest quarter (°C)	1310.25
Coefficient of variation of the annual rainfall (%)	2302.15
Mean monthly rainfall – winter ¹ months (mm)	11321.21
Mean monthly rainfall – summer ¹ months (mm)	874.11
Potential evaporation ² – July (mm)	1042.05
Potential evaporation ² – January (mm)	1650.52
Mean annual A-pan evaporation (mm) ³	2186.92
Mean duration of the growing season (days)	2429.67
Moisture stress days ^{3,4} – July (%)	10075.72
Moisture stress days ⁴ – January (%)	1026.62
Mean soil moisture days ^{3,5} – summer ⁶ (days)	848.76
Mean soil moisture days ^{3,5} – winter ⁶ (days)	10042.60
Mean winter ¹ heat units ³ (degree-days > 10 °C)	515.89
Mean summer ¹ heat units ³ (degree-days > 10 °C)	838.32
Annual heat units ³ (degree days > 18 °C)	785.31

¹Winter = April to September, summer = October to March.

²Estimated using the FAO58 method (Schulze et al., 1997).

³Variables used by Rutherford et al. (2000).

⁴Soil moisture stress days are those where transpiration is < 50% of the potential maximum because plant available water in the A-horizon was < 40% of the maximum (Schulze *et al.*, 1997).

⁵The inverse of the soil moisture stress days summed for the respective months.

⁶Winter = May to August, summer = November to February.

Climate data and variables

The basic climatic data were taken from Schulze *et al.* (1997) and are available for a $1' \times 1'$ latitude × longitude grid (Table 3). The seasonal rainfall concentration values, calculated using the method due to Markham (1970), were taken from Schulze *et al.* (1997). Fifteen basic climatic parameters were derived by Mgidi *et al.* (2007) from climatic data in Schulze *et al.* (1997) and were

included in this analysis. We also included the variables used by Rutherford *et al.* (2000) which were based on those derived by Schulze & Perks (1999) for current climatic conditions.

Selection of variables and thresholds

It is very difficult to be sure whether the limits of a distribution are determined only by environmental or biotic factors or by some combination of these distinct types. In this study climatic factors were chosen because they tend to define absolute limits which are easy to interpret and easy to use for delimiting potential distributions, particularly temperature and moisture availability (see review by Woodward, 1987; Cramer et al., 2001). Non-climatic factors which have been used in studies of the Cape flora include vegetation indices, elevation (which is correlated with temperature and rainfall) and soil fertility, texture and pH, which have been found to have statistically significant associations with the distributions of fynbos species (Rutherford et al., 2000; Gelfand et al., 2006; Latimer et al., 2006; Midgley et al., 2006). However, a preliminary screening indicated that soil fertility (which is correlated with soil texture and pH) is only weakly associated with the presence and absence of Hakea species, so soil fertility was not included in this analysis. The robustness of the chosen limits depends on the reliability of the species records and the climate data interpolation. The resolution of the climatic data and the Protea Atlas data are the best available.

The initial screening of the climatic variables was done using a standard chi-square test of the goodness of fit between values for cells with Hakea and all the cells in the study area. The number of classes and class intervals were chosen to minimize the occurrence of zero counts in a class. Variables with non-significant chi-square values were excluded from further analysis. The variables used to define the potential distribution were also assessed visually, focusing on areas where there are steep climatic gradients and Hakea invasions are known to have been present for several decades in the vicinity (within 1-2 km). The area chosen for this was the mountain areas inland of Cape Town where Hakea was introduced more than 100 years ago and historical invasions were very extensive (Fugler, 1979). A three-step process was followed: (1) A threshold value was selected for a given climatic variable and overlaid on the presence records using GIS. (2) The value was adjusted iteratively and the boundary redrawn until it enclosed all the presence records while still including potentially invadable areas. The latter criterion was included because most of the lowland areas, with lower rainfall, have been converted for agriculture (Cowling & Pressey, 2003) and are no longer invadable (e.g. cultivated land). Thus the lower threshold for rainfall had to take land conversion into account. The invadable areas were taken as being the natural remnants in the land-cover data set created for conservation planning (C.A.P.E. Natural Remnants http://bgis.sanbi.org/c.a.p.e/ naturalRemnants.asp accessed February 2007). (3) The selected threshold value was tested across all the presence records and, in most cases, covered the rest of the records as well. There were just 14 exceptions to the threshold value for the mean minimum temperature of the coldest month, which were all located in one

portion of a remote mountain range. Adjustment of the threshold by 2 °C to include these outliers would add 15% to the total area of the envelope so it was more reasonable to accept that the modelled climate parameters for this remote area were inaccurate than to adjust a threshold which was robust in well-instrumented areas.

Spatial autocorrelation and multicollinearity

Spatial autocorrelations and multicollinearity within and between the input variables alter the model statistics and bias the goodness of fit measures, making them overestimate the accuracy of the model (Legendre, 1993; Guisan & Zimmermann, 2000; Robertson *et al.*, 2001). Multi-collinearity was minimized by a priori exclusion of correlated variables and spatial correlation was reduced by randomly subsampling the climate data. This does not eliminate spatial autocorrelation completely (Legendre, 1993; Guisan & Thuiller, 2005; Dormann, 2007). The remaining spatial autocorrelation is not considered to be an issue in this analysis because the emphasis is on comparing model outputs, all of which would be subject to the same degree of bias, rather than the accuracy with which specific models fitted the data.

HAKEA SPECIES RESULTS

Abundance data

Both the complete and the random sample presence data sets on population sizes and on the numbers of records per 1MS cell were tested to see if they were related to any climatic variables. None of the relationships were significant nor were there significant differences between abundance classes in the mean values of any of the climatic values. Thus there was no evidence that the number of plants in the population or the number of populations per 1MS was greater near the centre of the range of values for different climatic variables. Nor was there any evidence that small populations or numbers or records per unit area occur mainly at or near the margins of the distributions. Therefore the data used in the subsequent analyses were treated as simple presence–absence values.

Climatic variables

The best predictors were, in order: rainfall in the coldest quarter (RCQ), mean monthly rainfall of the winter months, percentage of moisture stress days in July and mean soil moisture days in winter (Table 3). The next highest chi-square values were for: rainfall in the driest quarter (RDQ) and the rainfall in the driest month. The RCQ has a strong and nonlinear relationship with the occurrence of *Hakea* (Fig. 2). The high chi-square value (Table 3) is mainly because only 15% of the *Hakea*-invaded 1MS cells have an RCQ of < 100 mm compared with 95% of those in the study area. The corresponding values for < 50 mm are 0.7 and 89%, respectively. Inspection of the spatial distribution showed that the low frequency of *Hakea* records from areas with > 1000 mm is because most of these 1MS cells are found at high altitude



Figure 2 Relative frequency of $1' \times 1'$ squares (1MS) in different classes of rainfall during the coldest quarter (mm) for South Africa and for those invaded by *Hakea* species. Values on the *x*-axis are the upper boundary of the class. Data from Mgidi *et al.* (2007), raw data from Schulze *et al.* (1997). Note the logarithmic scale on the *y*-axis and the truncated bar for < 100 mm.



Figure 3 Relative frequency of $1' \times 1'$ squares (1MS) in different classes of rainfall of the driest quarter (mm) for South Africa and for those invaded by *Hakea* species. Values on the *x*-axis are the upper boundary of the class. Data from Mgidi *et al.* (2007), raw data from Schulze *et al.* (1997). Note the logarithmic scale on the *y*-axis.

where the limiting factors may be low winter minimum temperatures rather than rainfall.

The RDQ also shows a markedly different frequency distribution than that for South Africa, with no *Hakea* invasions at < 5 mm and very few at less than 10 mm (Fig. 3). The rainfall concentration had a relatively low chi-square value. The variables moisture stress days and mean soil moisture days showed marked seasonal contrasts: both having high chi-square values for winter months (July and May to August, respectively) and low values for the summer months (Table 3). The differences for the winter months seem to be primarily due to most *Hakea* records being from the winter rainfall areas in the west and south of their range. The high chi-square value for the percentage of moisture stress days in July is mainly because there are only nine *Hakea*-invaded cells with > 75% of days compared with about 50% of the study area.



Figure 4 Summary of the relative frequency all $1' \times 1'$ grid cells in the study area and for those invaded by *Hakea* in classes of the mean minimum temperature for the coldest month (May to August). Data from Mgidi *et al.* (2007), raw data from Schulze *et al.* (1997). Note the logarithmic scale on the *y*-axis.

None of temperature variables had comparably high chi-square values (Table 3). The highest chi-square value was for the mean temperature of the wettest quarter followed by the average temperature range and the temperature of the coldest month (TCM). There were no recorded *Hakea* invasions with a TCM of ≤ 0 °C (Fig. 4) with only one *Hakea* invaded 1MS occurring in an area with an estimated minimum of < 1 °C. Nearly 30% of all the 1MSs in the South Africa region have a TCM of ≤ 1 °C. None of the variables using heat units had high chi-square values.

All the recorded populations of *Hakea* were confined to elevations less than 1400 m. Only about 5% of the 1MSs in the study area have a greater elevation, but it is notable that these areas are uninvaded even where they are adjacent to numerous *Hakea* populations and well within the seed dispersal range.

Defining the climatic thresholds for screening pseudo-absence records

The final choice of the variables and the thresholds that would be used to define the boundaries of the potential distribution of *Hakea* was made by: (1) Selecting the variables with high chi-square values and comparing them visually with the distribution of the *Hakea* records from the Protea Atlas data set generalized to a 1MS resolution. (2) Selecting the variable with the next highest chi-square that was not strongly correlated with the RCQ.

The strongest candidates were: mean winter monthly rainfall (MWMR), moisture stress days in July and mean soil moisture days in winter and rainfall of the driest quarter (RDQ). The first three were all strongly correlated ($r^2 > 0.75$, P < 0.01). Soil moisture balance modelling introduces additional sources of error so the simplest variables, those derived directly from measurements, were chosen. The final choice was to use the RCQ as the first variable. The next uncorrelated variable with a high chi-square value was the rainfall of the driest quarter (RDQ). After some preliminary analysis and visual interpretation the threshold for the RCQ was set at 30 mm and the RDQ at 6 mm which excluded only 14 (0.7%) of the 2065 *Hakea*-invaded 1MS grid cells. The addition of further variables made little difference to the boundaries so no further variables were chosen for use in screening.

Modelling with and without predetermined climatic thresholds

The SRE model was used to generate a map which would show the climate envelope defined for the *Hakea* species using all the presence records from the Protea Atlas data (Fig. 5) and all the climate variables: rainfall of the coolest quarter (RCQ), rainfall of the driest quarter (RDQ), mean minimum temperature of the



Figure 5 The potential distribution of *Hakea* species using the surface range envelope (SRE) model (0 = absence and 1 = presence) based on all the Protea Atlas records. QDS = the quarter degree squares where *Hakea* populations were recorded during the Protea Atlas study (data supplied by Dr A. G. Rebelo, SANBI).

© 2008 The Authors



Figure 6 The potential distribution of *Hakea* species using the surface range envelope (SRE) model: 0 = absence and 1 = presence based on a subsample of the Protea Atlas records. QDS = the quarter degree squares where *Hakea* populations were recorded during the Protea Atlas study (data supplied by Dr A. G. Rebelo, SANBI).

coldest month (TCM), mean rainfall of the summer months (MSMR), mean duration of the growing season (MDGS, days) and mean temperature of the wettest quarter (TWQ). The map shows the full extent of the potential distribution for all species based on the maximum and minimum values of all the climatic variables for all the presence records using a scale of either absent (0) or present (1). The potentially invadable area includes all the coastal and inland mountain ranges of the Western and Eastern Cape and extends in a broad band up the east coast but not into north-eastern KwaZulu-Natal. The results of the SRE model fitted to the subsample of the Protea Atlas records used in the other models (below) shows a much more restricted climate envelope, particularly in the Eastern Cape and KwaZulu-Natal (Fig. 6) although it still includes virtually all the presence records.

The GLM model for all the species, and with all the pseudo absence records, used the two threshold variables (RCQ and RDQ and the TCM (Table 4). All the fitted relationships were nonlinear and used third-order (cubic) polynomials. Both RCQ and RDQ had regressions with a threshold near the chosen values (30 mm and 6 mm, respectively) and TCM used a quadratic form with a maximum probability at about 4 °C and zero probability at 0 °C and 10 °C, similar to the actual distribution (Fig. 4).

The potential distribution of *H. sericea* based on a GLM using all the selected climatic variables and all the absences is shown in Fig. 7. The climate variables used for this model were RCQ, RDQ, MSMR, MDGS and TWQ (Table 4). The GLM model used polynomial models with RCQ, RDQ and TCM as variables. The selected threshold to transform the probability values into binary presence and absence is 0.343 and this boundary would include 94% of the sample records. This still excludes some of the records in the 10% random sample of presence records, particularly the record from the north-eastern part of the Eastern Cape. The potential distribution based on a priori exclusion of false absence records with an RCQ < 30 mm and RDQ < 6 mm shows some interesting contrasts (Fig. 8). The GLM used TWQ, MSMR and TCM as the strongest predictors (Table 4). The extent of the potential distribution has been increased and the boundaries are now much sharper with few areas in intermediate probability classes. The threshold value for presence–absence is 0.56, which is higher than for the previous analysis and includes 99.8% of the sampled records but still excludes the presence records in the Wild Coast. The modelled potential distribution in this case includes montane areas on the northern side of Lesotho as well as areas of the northern coast of KwaZulu-Natal (Fig. 8). These areas were not included in the SRE model for the full set of *Hakea* records (Fig. 5) or for the 10% random subsample (Fig. 6).

Similar results were obtained for H. drupaceae and H. gibbosa (Fig. 9). In each case, excluding the false absences (i.e. those with a high likelihood of being true presences) increased the extent of the area projected by the model as indicating a high invasion potential and the proportion of intermediate probabilities declined. Taking H. sericea as an example, 1.5% of the 1MS grid cells were classified as having a presence probability of > 0.8 for the full pseudo-absence set compared with 6.5% when the false absences were excluded. The zero probability 1MS cells increased from 90.0% to 92.7% in H. sericea and H. drupaceae but decreased in H. gibbosa. The predicted distributions included most of the recorded populations of these two species with a high probability of their being present. Using the full sample of pseudo-absence data for H. sericea, only 4.5% of the presence records were given an invasion potential probability of less than the threshold of 0.34 (Table 4), and only 3.4% of the pseudo-absence records a probability > 0.34. When the false absences were excluded, these percentages decreased to 0.75% and 0.22%, respectively, although the presence-absence probability threshold was now 0.56.

Species	Excluding false absence records	Variable	Function	Null deviance	Degrees of freedom	Residual deviance	AIC	Cut-off (ROC)	Sensitivity	Specificity
All	No	RCQ	3rd order	1553.28	1385	231.36	251.36	0.412	96.53	96.44
		RDQ	3rd order							
		TCM	3rd order							
	Yes	TWQ	2nd order	1503.67	1285	13.13	12.2	0.04	99.18	99.26
		MSMR	3rd order							
		TCM	2nd order							
H. drupacea	No	RCQ	2nd order	302.02	1078	55.64	65.64	0.08	98.04	97.52
		TCM	2nd order							
	Yes	TWQ	2nd order	308.48	978	< 0.01	12.0	0.99	100	100
		MSMR	2nd order							
		TCM	1st order							
H. gibbosa	No	MSMR	2nd order	406.42	1095	33.44	49.45	0.001	97.33	98.06
		RCQ	3rd order							
		PEt07	2nd order							
	Yes	TWQ	2nd order	442.38	995	0.001	10.0	0.99	100	100
		MSMR	2nd order							
H. sericea	No	RCQ	3rd order	1358.20	1323	221.78	239.78	0.34	96.26	96.11
		RDQ	3rd order							
		TCM	2nd order							
	Yes	TWQ	2nd order	1340.31	1223	14.81	38.81	0.56	99.75	99.77
		MSMR	3rd order							
		TCM	3rd order							
		MDGS	3rd order							

Table 4 Summary of the variables and the functions fitted to the data using the generalized linear model with summary statistics on the fit ofthe model. The second- and third-order functions are quadratic and cubic, respectively.

AIC, Akaike information criterion; RCQ, rainfall of the coldest quarter; RDQ, rainfall of the driest quarter; TCM, temperature of the coldest month; TWQ, temperature of the wettest quarter; MSMR, mean summer month (October to March) rainfall; PEt07, July potential evaporation; MDGS, mean duration of the growing season in days; Cut-off (ROC), the probability threshold for presence as determined from the area under the receiver operating characteristic curve; sensitivity, the proportion of the presence records correctly predicted; specificity, the proportion of absence records correctly predicted.



Figure 7 The potential distribution of *Hakea sericea* using a generalized linear model (GLM) shown as a probability surface fitted to the full set of climatic variables. QDS = the quarter degree squares where *H. sericea* populations were recorded during the Protea Atlas study (data supplied by Dr A. G. Rebelo, SANBI). This figure is available at high resolution in the Supplementary Material (Figure S1).



Figure 8 The potential distribution of *Hakea sericea* using a generalized linear model (GLM) shown as a probability surface fitted to the reduced set of climate variables and excluding absence samples within the predefined climate boundaries. See the text for more information. QDS = the quarter degree squares where *H. sericea* populations were recorded during the Protea Atlas study (data supplied by Dr A. G. Rebelo, SANBI). This figure is available at high resolution in the Supplementary Material (Figure S2).

DISCUSSION

The results of this study show that a priori screening of absence records using climatic variables can be a useful approach to refining our understanding and ability to define the potential distribution of an invading species. The chi-square analysis, supplemented with visual assessment of an overlay of the distribution records on the climatic variables, proved to be a useful way of identifying the climatic variables with strong relationships with the presence and absence of *Hakea* species. One advantage of the chi-square analysis is that it does not assume a particular form of environmental response, meeting the concerns of Austin (1980, 1999, 2007) and others (e.g. Green, 1971) about fitting appropriate forms of the response functions. Where the test shows there are marked differences, and these are situated in the upper or lower range of the values of a given variable, they give a strong indication of the existence of a threshold.

The choice of variables selected for the GLM and GAM models was also consistent with the chi-square ranking (Table 4): RCO, RDQ and TCM were generally included when the full pseudoabsence data set was used. When the RCQ and RDQ were excluded the TCM, MSMR and TWQ were consistently used. The consistent use of these variables suggests that they are robust variables to use in screening plants and should be considered for other studies in the future as they can be derived from very basic climatic data. The importance of TCM is interesting, particularly the possibility that Hakea species have been prevented from invading mountain areas with TCM less than about 1 °C which corresponds roughly with an elevation greater than 1400 m. This relationship was not identified in a previous study of the factors determining its distribution (Richardson, 1984) although TCM was used to define H. sericea's invasion potential by Richardson et al. (2000). The mapped native distributions of these species in Australia do not include high mountain areas (Flora of Australia Online, http://www.environment.gov.au/biodiversity/abrs/ online-resources/flora/main/, accessed 13 November 2006), indicating that these species may be sensitive to low temperatures.

Our results emphasize that relationships between environmental factors and species responses are often nonlinear and asymmetric (Austin, 1980). In this study most of the relationships were curvilinear, with many having a sigmoid or asymptotic form (often represented by a third-order polynomial) or quadratic (Table 4), although in some cases they seemed to be multimodal. The issue of non-unimodal relationships between environmental factors and species responses is an interesting one and difficult to conceptualize. Intuitively, we would expect a unimodal response to a single factor, an approach exemplified in the results of the classical gradient studies of Whittaker (1975). A multimodal relationship with a particular variable could be the result of the influence of a second, probably uncorrelated, climatic or non-climatic variable which results in there being relatively few occurrences within a particular range of the first variable. Deviations from monotonic relationships may offer new insights and merit some additional research. An alternative approach to the problem of defining the a priori thresholds more objectively is quantile regression, which can fit upper and lower quantiles to data similar to those used in this study (Koenker & Hallock, 2001; Cade & Noon, 2003). Algorithms and R-code for quantile regression using binomial response variables are currently under development (R. Koenker, personal communication, February 2007).

Correlation analyses found that many of the input variables were strongly correlated with each other, both positively and negatively. In this study we deliberately excluded the strongly correlated variables because we believe that the a priori exclusion of correlated variables is an important step in reducing multicollinearity in the independent variables. We suggest that the use of the basic variables such as rainfall and temperature reduces the uncertainties in the estimates and is just as meaningful ecologically. A number of other studies have included minimum temperature variables (e.g. TCM) and elevation (e.g. Gelfand *et al.*, 2005;



Figure 9 The potential distributions of *Hakea drupacea* and *Hakea gibbosa* using a generalized linear model (GLM) shown as probability surfaces fitted to the full set of climate variables (a and b) and excluding pseudo-absence samples with the predefined climate boundaries (c and d). See the text for more information. QDS = the quarter degree squares where *H. drupacea* or *H. gibbosa* populations were recorded for the Protea Atlas (data supplied by Dr A. G. Rebelo, SANBI). This figure is available at high resolution in the Supplementary Material (Figure S3).

Latimer *et al.*, 2006) but the strong correlations between these variables imply that using both increases the multicollinearity.

The lack of a strong relationship between the climatic factors included in this analysis and the number of populations per 1MS or the population size, from the Protea Atlas data, was unexpected. This was particularly so because of the significance of the chi-square analyses of the relationships between the values of the climatic variables for all records and those where Hakea had invaded (Table 3, Figs 2-4). A previous study of Protea species found that abundance data were useful in understanding environmental preferences and site history (Latimer et al., 2006). Our expectation was that the same conditions that favour the establishment of populations would also favour the competitiveness and growth rates of the species and the recruitment of large populations, as recognized by Grinnell (1917) (for recent reviews and discussions of this issue see Arim et al., 2006; Austin, 2007), although non-climatic factors may also determine abundance (Hutchinson, 1957; Barry & Elith, 2006). The typically exponential relationship between the time since that population was founded and number of individuals in that population could be a confounding factor (see Wilson et al., 2007), provided that new and old invasion foci were more or less evenly distributed or that old foci occurred near the edge of the current distribution. The impact of human activities in determining where the source populations were established may be a factor. Livestock corrals tend to be situated where the climatic conditions are suitable for human habitation. This would tend to be on the lower mountain slopes and in valleys where the climatic conditions are closer to those that indicate the limits of the current distributions of the three species.

The robustness of this absence filtering approach depends strongly on an adequate knowledge of the ecology, invasion history and current distribution of that species but it also provides more insights than a single 'best' model (Lütolf *et al.*, 2006). The more knowledge and insight that can be generated through the modelling process, the better the guidance that can be provided to managers. Given that agencies involved with invasive species have to make choices anyway, studies that can reduce the degree of uncertainty or offer new insights can provide valuable information for designing control strategies and making tactical choices.

ACKNOWLEDGEMENTS

This study was supported by the Strategic Research Panel of the CSIR. The Protea Atlas data base records for *Hakea* species were provided by Dr A. G. Rebelo of the South African National Biodiversity Institute. The SAPIA data base has been developed and maintained by the Plant Protection Research Institute of the South African Agricultural Research Council. We thank Lesley Henderson for permission to use the data. This study contributes to the International Research Network (GDRI) project 'France South Africa – dynamics of biodiversity in Southern African ecosystems and sustainable use in the context of global change: processes and mechanisms involved'. W.T. was partly funded by the EU FP6 MACIS Specific Targeted Project (Minimization of and adaptation to climate change: impacts on biodiversity, no. 044399).

REFERENCES

- Arim, M., Abades, S.R., Neill, P.E., Lima, M. & Marquet, P.A. (2006) Spread dynamics of invasive species. *Proceedings of the National Academy of Sciences USA*, **103**, 374–378.
- Austin, M.P. (1980) Searching for a model for use in vegetation analysis. Vegetatio, 42, 11–21.
- Austin, M.P. (1999) A silent clash of paradigms: some inconsistencies in community ecology. Oikos, 86, 170–178.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Austin, M. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.
- Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43, 413–423.
- Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.
- Cade, B.S. & Noon, B.R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1, 412–420.
- Callaway, R.M. & Maron, J.L. (2006) What have exotic plant invasions taught us over the past 20 years? *Trends in Ecology and Evolution*, **21**, 369–374.
- Cowling, R.M. & Pressey, R.L. (2003) Introduction to systematic conservation planning in the Cape Floristic Region. *Biological Conservation*, **112**, 1–13.
- Cramer, W., Bondeau, A., Woodward, F.I., Prentice, C.I., Betts, R.A., Brovkin, V., Cox, P.M., Fisher, V., Foley, J.A., Friend, A.D., Kucharik, C., Lomas, M.R., Ramankutty, N., Sitch, S., Smith, B., White, A. & Young-Molling, C. (2001) Global response of terrestrial ecosystem structure and function to CO₂ and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357–373.
- Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129–138.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151.
- Fugler, S.R. (1979) Some aspects of the autecology of three Hakea species in the Cape Province, South Africa. Unpublished MSc Thesis, University of Cape Town.
- Gelfand, A.E., Latimer, A., Wu, S. & Silander, J.A., Jr (2005) Building statistical models to analyze species distributions. *Hierarchical modelling for the environmental sciences: statistical methods and applications* (ed. by J.A. Clark and A.E. Gelfand), pp. 77–97. Oxford University Press, Oxford, UK.

- Gelfand, A.E., Silander, J.A., Jr, Wu, S., Latimer, A., Lewis, P.O., Rebelo, A.G. & Holder, M. (2006) Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, 1, 41–92.
- Green, R.H. (1971) A multivariate statistical approach to the Hutchinsonian niche: bivalve molluscs of central Canada. *Ecology*, **52**, 543–556.
- Grinnell, J. (1917) Field tests of theories concerning distributional control. *The American Naturalist*, **51**, 115–128.
- Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.McC., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43**, 386–392.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Heikkinen, R.K., Luoto, M., Araújo, Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 1–27.
- Henderson, L. (1998) Southern African Plant Invaders Atlas (SAPIA). *Applied Plant Science*, **12**, 31–32.
- Hierro, J.L., Maron, J.L. & Callaway, R.M. (2005) A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. *Journal of Ecology*, **93**, 5–15.
- Higgins, S.I., Clark, J.S., Nathan, R., Hovestadt, T., Schurr, F., Fragoso, J.M.V., Aguiar, M.R., Ribbens, E. & Lavorel, S. (2003) Forecasting plant migration rates: managing uncertainty for risk assessment. *Journal of Ecology*, **91**, 341–347.
- Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecologicalniche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.
- Hirzel, A.H., Helfer, V. & Metral, F. (2001) Assessing habitat suitability models with a virtual species. *Ecological Modelling*, **145**, 111–121.
- Hoffmann, J.H. & Moran, V.C. (1998) The population dynamics of an introduced tree, *Sesbania punicea*, in South Africa, in response to long-term damage caused by different combinations of three species of biological control agents. *Oecologia*, **114**, 343–348.
- Hooge, P.N. & Eichenlaub, B. (1997) *Animal movement extension to ArcView*, Version 1.1. Alaska Biological Science Center, US Geological Survey, Anchorage, AK, USA.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
- Joshi, K. & Vrieling, K. (2005) The enemy release and EICA hypothesis revisited: incorporating the fundamental difference between specialist and generalist herbivores. *Ecology Letters*, **8**, 704–714.
- Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13**, 853–867.

- Keane, R.M. & Crawley, M.J. (2002) Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology and Evolution*, 17, 164–170.
- Koenker, R. & Hallock, K.F. (2001) Quantile regression. *Journal* of *Economic Perspectives*, **15**, 143–156.
- Kruger, F.J., Richardson, D.M. & van Wilgen, B.W. (1986)
 Processes of invasion by alien plants. *The ecology and management of biological invasions in South Africa* (ed. by I.A.W. Macdonald, F.J. Kruger and A.A. Ferrar), pp. 144–155. Oxford University Press, Cape Town.
- Latimer, A., Wu, S., Gelfand, A.E. & Silander, J.A. (2006) Building statistical models to analyze species distributions. *Ecological Applications*, **16**, 33–50.
- Lavergne, S. & Molofsky, J. (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences USA*, 104, 3883–3888.
- Le Maitre, D.C., Krug, R.M., Hoffmann, J.H., Gordon, A.J. & Mgidi, T.N. (2008) *Hakea sericea*: development of a model of the impacts of biocontrol on population dynamics and rates of spread of an invasive species. *Ecological Modelling*, **212**, 342–358.
- Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659–1673.
- Lütolf, M., Kienast, F. & Guisan, A. (2006) The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology*, **43**, 802–815.
- McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.
- Markham, C.G. (1970) Seasonality of precipitation in the United States. *Annals of the Association of American Geographers*, **60**, 593–597.
- Mau-Crimmins, T., Schussman, H.R. & Geiger, E.L. (2006) Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling*, **193**, 736–746.
- Mgidi, T.N., Le Maitre, D.C., Schonegevel, L., Nel, J., Rouget, M. & Richardson, D.M. (2007) Alien plant invasions – incorporating emerging invaders in regional prioritization: a pragmatic approach for Southern Africa. *Journal of Environmental Management*, **84**, 173–187.
- Midgley, G.F., Hughes, G.O., Thuiller, W. & Rebelo, A.G. (2006) Migration rate limitations on climate change-induced range shifts in Cape Proteaceae. *Diversity and Distributions*, **12**, 555–562.
- Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W. & Levin. S.A. (2002) Mechanisms of long-distance dispersal of seeds by wind. *Nature*, 418, 409–413.
- Neser, S. (1983a) Rock hakea (*Hakea gibbosa* (Sm.) Cav.). *Plant invaders: beautiful, but dangerous*, 3rd edn (ed. by C.H. Stirton), pp. 72–75. Department of Nature and Environmental Conservation, Cape Town.
- Neser, S. (1983b) Sweet hakea (Hakea suaveolens R.Br.). Plant invaders: beautiful, but dangerous, 3rd edn (ed. by C.H.

Stirton), pp. 80–83. Department of Nature and Environmental Conservation, Cape Town.

Neser, S. & Fugler, S.R. (1983) Silky hakea (*Hakea sericea*. Schrad.). *Plant invaders: beautiful, but dangerous*, 3rd edn (ed. by C.H. Stirton), pp. 76–79. Department of Nature and Environmental Conservation, Cape Town.

Parker, J.D., Burkepile, D.E. & Hay, M.E. (2006) Opposing effects of native and exotic herbivores on plant invasions. *Science*, **311**, 1459–1461.

- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.E. & Lees, D.C. (2006) Model-based uncertainty in species' range prediction. *Journal of Biogeography*, 33, 1704–1711.
- Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology*, **78**, 419–433.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Pyšek, P. & Hulme, P.E. (2005) Spatio-temporal dynamics of plant invasions: linking pattern to process. *Ecoscience*, **12**, 302–315.
- Rebelo, A.G. (1991) *Protea Atlas manual: instruction booklet to the Protea Atlas Project.* Protea Atlas Project, South African. Botanical Institute, Cape Town.
- Reese, G.C., Wilson, K.R., Hoeting, J.A. & Flather, C.H. (2005) Factors affecting species distribution predictions: a simulation modelling experiment. *Ecological Applications*, **15**, 554– 564.
- Richardson, D.M. (1984) A cartographic analysis of physiographic features influencing the distribution of *Hakea* spp. in the south-western Cape. *South African Forestry Journal*, **128**, 36–40.

Richardson, D.M. & Bond, W.J. (1991) Determinants of plant distribution: evidence from pine invasions. *The American Naturalist*, **137**, 639–668.

Richardson, D.M. & Thuiller, W. (2007) Home away from home – objective mapping of high-risk source areas for plant introductions. *Diversity and Distributions*, **13**, 299–312.

Richardson, D.M., van Wilgen, B.W. & Mitchell, D.T. (1987) Aspects of the reproductive ecology of four Australian *Hakea* species (Proteaceae) in South Africa. *Oecologia*, **71**, 345–354.

Richardson, D.M., Bond, W.J., Dean, W.R.J., Higgins, S.I., Midgley, G.F., Milton, S.J., Powrie, L.W., Rutherford, M.C., Samways, M.J. & Schulze, R.E. (2000) Invasive alien species and global change: a South African perspective. *Invasive species in a changing world* (ed. by H.A. Mooney and R.J. Hobbs), pp. 303–349. Island Press, Washington DC.

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

Rutherford, M.C., Midgley, G.F., Bond, W.J., Powrie, L.W., Roberts, R. & Allsopp, J. (2000) Plant biodiversity. *Climate change impacts in Southern Africa* (ed. G. Kiker), pp. 59.
Report to the National Climate Change Committee, Department of Environment Affairs and Tourism, Pretoria.

- SAS Institute Inc. (2004) *SAS 9.1.2.* SAS Institute Inc., Cary, NC, USA.
- Schulze, R.E. & Perks, L.A. (1999) Assessment of the impact of climate change. Final Report to the South African Country Studies Climate Change Programme, School of Bioresources Engineering and Environmental Hydrology, University of Natal, Pietermaritzburg, South Africa.
- Schulze, R.E., Maharaj, M., Lynch, S.D., Howe, B.J. & Melvil-Thomson, B. (1997) *South African atlas of agrohydrology and climatology*. Water Research Commission, Pretoria.
- Shaughnessy, G.L. (1986) A case study of some woody plant introductions to the Cape Town area. *The ecology and management of biological invasions in Southern Africa* (ed. by I.A.W. Macdonald, F.J. Kruger and A.A. Ferrar), pp. 271–308. Oxford University Press, Cape Town.
- Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1–13.
- Taylor, C.M. & Hastings, A. (2005) Allee effects in biological invasions. *Ecology Letters*, **8**, 895–908.
- Thuiller, W. (2003) BIOMOD: optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, **9**, 1353–1362.
- Thuiller, W., Richardson, D.M., Pyšek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.
- Thuiller, W., Richardson, D.M., Rouget, M. & Proches, Ş. & Wilson, J.R.U. (2006) Interactions between environment, species traits, and human uses describe patterns of plant invasions. *Ecology*, **87**, 1755–1769.
- Vilà, M. & Weiner, J. (2004) Are invasive plant species better competitors than native plant species? evidence from pair-wise experiments. *Oikos*, **105**, 229–238.
- Welk, E. (2004) Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling*, **179**, 551–567.
- White, E.M., Wilson, J.C. & Clarke, A.R. (2006) Biotic indirect effects: a neglected concept in invasion biology. *Diversity and Distributions*, **12**, 443–455.
- Whittaker, R. (1975) *Communities and ecosystems*. Macmillan, New York.
- Wilson, J.R.U., Richardson, D.M., Rouget, M., Procheş, Ş., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions*, 13, 11–22.
- Woodward, F.I. (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge, UK.
- van Wyk, A.E. & Smith, G.F. (2001) *Regions of floristic endemism in Southern Africa*. Umdaus Press, Pretoria, South Africa.
- Zaniewski, A.E., Lehmann, A. & Overton, J.McC. (2000) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261–280.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Figure S1 The potential distribution of *Hakea sericea* using a generalized linear model (GLM) shown as a probability surface fitted to the full set of climatic variables. See legend to Fig. 7 for additional details.

Figure S2 The potential distribution of *Hakea sericea* using a generalized linear model (GLM) shown as a probability surface fitted to the reduced set of climate variables and excluding absence samples within the predefined climate boundaries. See legend to Fig. 8 for additional details.

Figure S3 The potential distributions of *Hakea drupacea* and *Hakea gibbosa* using a generalized linear model (GLM) shown as probability surfaces fitted to the full set of climate variables (a and b) and excluding pseudo-absence samples with the predefined climate boundaries (c and d). See legend to Fig. 9 for additional details.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/ j.1466-8238.2008.00407.x (This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BIOSKETCHES

David Le Maitre obtained his PhD from the University of Cape Town. He is principal researcher in the Biodiversity and Ecosystem Services Research Group, Natural Resources and the Environment, CSIR. His main interests include invasion ecology and assessing the impacts of invasive species on ecosystem functioning and services.

Wilfried Thuiller has a PhD from the University of Montpellier. Currently he is a researcher at the Centre National de la Recherche Scientifique in Grenoble. His main interest is to understand and assess the impacts of global change on species distributions, communities and ecosystem functioning.

Lucille Schonegevel trained as a mathematics teacher and obtained her MSc in geography at the University of Stellenbosch. She did her MSc on the modelling of invasive plant invasion and control using a geographical information system and is interested in GIS applications in environmental modelling.

Editor: Brad Murray