CHAPTER 6

Improving species' distribution model outputs using jackknife

procedures

Improving species' distribution model outputs using jackknife procedures

Smit I.P.J[1,2]., Erasmus B.F.N.[2,6], Kshatriya M.[2], Smit C.F.[3], Eardley C.[4], Chown S.L.[5], Van Jaarsveld A.S.[2,6]

[1] Department of Statistical Support, Technikon Pretoria, Private Bag X680, Pretoria 0002, South Africa

[2] Conservation Planning Unit, Department Zoology and Entomology, University of Pretoria, Pretoria, 0002

[3] Department of Statistics, University of Pretoria, Pretoria 0002, South Africa

[4] Agricultural Research Council (ARC), Plant Protection Research Institute, Private Bag X134, Pretoria 0001, South Africa

[5] Department of Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

[6] Centre for Environmental Studies, University of Pretoria, Pretoria 0002, South Africa

Corresponding author:

Prof A.S. van Jaarsveld, Faculty of Science, Private Bag X1, Stellenbosch, 7602, South Africa

Tel: (+27-21) 808-3071

Fax: (+27-21) 808-3680

E-mail: asvanj@sun.ac.za

For submission to: *Ecological Modelling*

ABSTRACT

Bioclimatic envelope modeling of species' distributions can be misleading if no measure of uncertainty is associated with the predicted probability of occurrence values. Here we demonstrate how jackknife procedures can be used to generate such measures of uncertainty in the case of the distribution of a carpenter bee (*Xylocopa senior*) in South Africa. The modelled probability of occurrence surface for *Xylocopa senior* is more meaningful when interpreted in the context of the variation underlying these predictions, i.e., when more confidence/weight is ascribed to predictions with limited variation. This procedure is potentially of considerable value for any approach based on interpolated species distributions, including survey planning, area selection, modelling of climate change effects and biogeography. Resampling techniques can also be used as a means to provide insights into model performance and for crossvalidation, without the need for data partitioning or the acquisition of additional data.

INTRODUCTION

Bioclimatic models are routinely used to predict and define the broad scale limits of species distributions (Lindenmayer *et al.* 1991). Thus, where the spatial distribution of a taxon is not fully defined, bioclimatic analyses allow researchers to describe the probable limits of distribution. (Lindenmayer *et al.* 1991; Brzeziecki *et al.* 1995; Lindsay *et al.* 1998). Moreover, such models can be used to evaluate the risks and benefits likely to arise for species under future climatic scenarios (Lindenmayer *et al.* 1991; Jeffree & Jeffree 1994; Huntley *et al.* 1995; Sykes & Prentice 1995; Jeffree & Jeffree 1996; Sykes *et al.* 1996; Hill *et al.* 1999). The underlying principle of bioclimatic modelling is to determine whether there is any relationship between the known distribution of a species and some climatic variable or variables. Popular modelling approaches that have been employed include generalised linear regression (e.g. logistic regression) (Margules & Stein 1989; Austin *et al.* 1990; Walker 1990; Osborne & Tigar 1992; Cary & Ullyett 1993; Hill *et al.* 1999), classification trees (Walker 1990) and discriminant function analysis (Caughley *et al.* 1987) (see Guisan & Zimmerman (2000) for a comprehensive review of predictive habitat distribution models).

The product of such modelling is usually a map with grid coverage that is used to present the model's prediction of the probable distribution of a species. The model makes an individual prediction for every grid cell. The predicted distribution maps can be either *binary maps* or *probability of occurrence surface maps*, depending on the model used (see Osborne & Tigar (1992) for a combination). In the binary map, each grid cell is assigned either a presence or an absence rating (e.g. Huntley *et al.* 1995; Lindsay *et al.* 1998; Hill *et al.* 1999). A limitation of these maps is that they do not distinguish between areas that are most suitable and those that are only marginally suitable. Probability surface maps overcome this problem by assigning a probability of occurrence to each grid cell (see for example Walker 1990; Beerling *et al.* 1995; Erasmus *et al.* 2000, Erasmus *et al.* 2002). The probability assigned to a specific cell is an indication of the suitability of climate (i.e. the predictors in the model) in that grid cell for the species (Robertson *et al.* 2001). Since a continuum of suitability values is generated, the relative suitability of areas (with regard to the variables included as predictors in the model) can be compared. Because probability surface maps communicate more information than binary maps, we focus on the former approach.

Irrespective of the modelling approach used for predicting the probability surface of a species, the result is a single predicted probability value for each grid cell. These probabilities could also be called *best estimates,* because the models base their calculations on all the available data. The predicted distribution is therefore presented as a geographical area divided into grid cells with a single probability (best estimate) of the suitability of that locality for a specific species (as determined by the selected predictor variables) (e.g. Fig. 1). Although the best estimate gives an indication of the probability of a grid cell being climatically suitable, the best estimate is still a value of a stochastic variable (new data will necessarily lead to a new best estimate) and it can be misleading if no indication of the *uncertainty* associated with that value is provided. In their review, Guisan & Zimmerman (2000) identify the lack of spatially explicit uncertainty measurements of predicted probabilities as a key area requiring further development. One way of attributing a useful uncertainty measurement is to estimate the variation underlying each stochastic best estimate. All else being equal, more confidence can be placed in a best estimate associated with less variation than a best estimate associated with substantial variation. However, probabilities of climate suitability are usually not presented together with estimates of uncertainty since probability surface models predict only a single probability of occurrence for each grid cell, using all the relevant data available.

Using a case study, the present paper makes a start at addressing Guissan & Zimmerman's (2000) concerns regarding the lack of a spatially explicit uncertainty measure for interpolated distributions and shows how jackknife re-sampling procedures can be used to estimate the variation around each probability of occurrence value. The manner in which such estimates of variation can improve the interpretation of a probability surface map generated by a bioclimatic model is shown, and the implications of this approach are outlined.

MATERIAL AND METHODS

Xylocopa senior Vachal distribution records

Collection records from the Plant Protection Research Institute of South Africa and the South African Museum were combined into a data set containing 27 records of a carpenter bee, *Xylocopa senior* (Fig. 1). The data set of *Xylocopa senior* was chosen because 27 records do
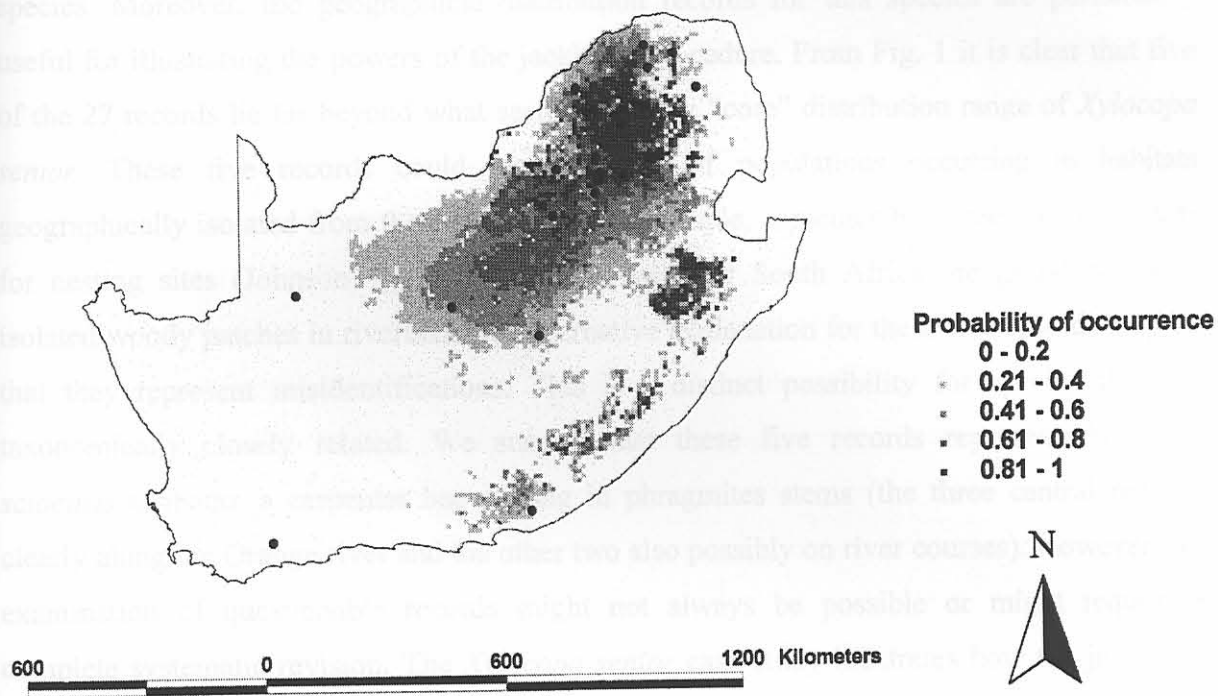
Fig. 1: Known distribution records (•) of *Xylopa senior* and the bioclimatically modelled probability of occurrence surface

not represent a comprehensive data set, and this is often a reality for poorly surveyed species, especially invertebrates (Koch *et al.* 2000). This example illustrates how limited distribution data can be used to provide useful information with regard to the probable distribution of a species. Moreover, the geographical distribution records for this species are particularly useful for illustrating the powers of the jackknife procedure. From Fig. 1 it is clear that five of the 27 records lie far beyond what seems to be the "core" distribution range of *Xylocopa senior*. These five records could be indicative of populations occurring in habitats geographically isolated from the core range. For example, carpenter bees need woody plants for nesting sites (Johnson 1997). The central parts of South Africa are grasslands with isolated woody patches in riverbeds. An alternative explanation for these five records may be that they represent misidentifications. This is a distinct possibility for species that are taxonomically closely related. We suspect that these five records represent *Xylocopa scioensis* Gribodo, a carpenter bee nesting in phragmites stems (the three central records clearly along the Orange river and the other two also possibly on river courses). However, re-examination of questionable records might not always be possible or might require a complete systematic revision. The *Xylocopa senior* case study illustrates how the jackknife procedure provides a way to further understand questionable or outlying distribution records.

Climate data

The Computing Centre for Water Research (CCWR) provided the historic climate data (thirty-year means; 1960-1990). These data comprise interpolated climate surfaces at a minute by minute resolution. These climate surfaces were resampled to a 10km x 10km grid cell size. Grid cell size was dictated by the total number of cells that can realistically be analysed given the available computer time without sacrificing output resolution too much. Bee point occurrence records were also generalised to this 10 km gridcell resolution.

The objective of this paper is not to make any concrete conclusions about the distribution of *Xylocopa senior* or the factors controlling its distribution, but rather to illustrate how the interpretation of a probability of occurrence surface of a species can be improved by employing jackknife procedures. Consequently, we did not strive to optimise the model by using the minimum number of significant variables, nor have we interpreted the model outcomes fully. The complexity of the model and the modeling approach is therefore considered relatively unimportant in this paper. Six climatic variables (mean temperature,

minimum temperature, maximum temperature, mean annual rainfall, mean rainfall for the driest month and mean rainfall for the wettest month) were selected as the climatic predictors. These variables were selected because they represent a suite of climatic variables likely to control the distribution of insects, either directly or indirectly (see Chown & Gaston 1999).

Model

The model we used here is a multivariate generalisation of the bivariate model described by Jeffree & Jeffree (1994). Briefly, using this method the grid cells where a particular species is recorded are referred to as *known* records. A scatter plot of the climate values of all grid cells for any two chosen climate variables is defined as the climate space. The climate values at points in the climate space that represent known records are used to construct the principle axes of an elliptical confidence region by using the sums of squares of the distance from the major and minor principal axis. The elliptical confidence region is superimposed on the climate space and can be interpreted as a climate envelope containing localities climatically similar to known records. All points falling within this climate envelope are then mapped back into geographical space to form an interpolated distribution of climatically similar areas where this species might potentially be found (see Jeffree & Jeffree (1994) for additional information). This approach by Jeffree & Jeffree (1994) is similar to the technique described by Sokal & Rohlf (1981) and relies heavily on the graphical interpretation of a two-dimensional scatter plot. In consequence, it has very limited application in scenarios where the climate space has to be multidimensional if three or more climate variables are needed to explain observed distribution patterns satisfactorily. Thus, we have adapted the original model to incorporate m climatic predictor variables. Not only is this a significant improvement on the original model, but it also allows the production of a probability surface of suitability for each species (Erasmus *et al.* 2000), rather than a more simplistic presence-absence distribution model. In the multivariate generalisation of the bivariate model the values of selected climate variables for each known record cell are plotted on an m-dimensional scatter plot (for the m climate variables), and mean climate values are subtracted to centre values around the origin of the multidimensional scatter plot. An $m \times m$ covariance matrix is calculated and this matrix is used as an input to calculate eigen values and eigen vectors for the covariance matrix. These eigen vectors from the orthogonal principle axis of an m-dimensional hyperspace with the origin representing the theoretical core of the species' distribution, as defined by the predictor climate variables (see Robertson *et al.* 2001 for a

detailed description of a similar, PCA-based model). The climate variable values of all grid cells are then transformed into this eigen vector space. These transformed climate variables are subsequently divided by the eigen values. In the resulting m-dimensional hyperspace, the distance of any particular unsampled grid cell from the origin represents a measure of the suitability of that locality for the specific species (Robertson *et al* 2001). The entire calculation for this multivariate technique is relatively straightforward and does not require considerable computing power. The outcome of this analysis is a probability surface map of suitability, as determined by the set of predictor variables, for a species. Hereafter, this probability surface of climate suitability will be referred to as the probability of occurrence.

Estimating variation using the jackknife procedure

By using jackknife procedures, n probability of occurrence values (n = size of data set) can be generated for each grid cell instead of a single value. This method re-calculates the statistic of interest, in this case a probability of occurrence, n times, each time using a different combination of n-1 of the data set records. The jackknife principle uses these n pseudovalues to estimate the variation associated with the statistic.

In the Erasmus *et al.* (2000) modelling procedure, the whole data set (n = 27) is used to calculate a single probability of occurrence for each grid cell (the best estimate, Fig. 1). The jackknife procedure re-calculates the probability of occurrence for each grid cell 27 (n) times using the same model with a different combination of 26 (n-1) records each time. These re-calculated estimates are referred to as the *jackknife pseudovalues*.

The jackknife pseudovalues can be used to calculate the standard error of the best estimate, which is an estimate of the variation associated with that best estimate, using the following formula (Efron & Tibshirani 1993; Shao & Tu 1995):

$$se_{jack} = \sqrt{n/(n-1) \times (\sum_{i=1}^{n} \hat{\theta}_i^2 - (\sum_{i=1}^{n} \hat{\theta}_i)^2 / n)} \qquad \text{(eqn 1)}$$

where se_{jack} = estimated standard error of the best estimate;

   $\hat{\theta}_i$ = i th jackknife pseudovalue;

n = size of data set.

se_{jack} is an estimate of the standard error of the best estimate and does not merely represent the standard deviation of the n jackknife pseudovalues.

The se_{jack} and the coefficient of variation ($\dfrac{se_{jack}}{best\ estimate}$) are therefore measures of variation associated with the best estimate. The se_{jack} is a measure of *absolute* variation while the coefficient of variation is a measure of *relative* variation (variation relative to the best estimate) (see results section). In reality, by assessing the best estimate's variation, the jackknife procedure has generated additional information from the original data set. These variance estimates are likely to improve the interpretation of the generated probability surfaces considerably.

RESULTS

Best estimate probability surface (model prediction)

The probability of occurrence surface (best estimates) for *Xylocopa senior* in South Africa, using a suite of six climatic variables as predictors in the model (Erasmus *et al.* 2000), is illustrated in Fig. 1. The model's prediction seems to provide a reasonably good fit for the data records, except for the distortion caused by five influential records which "pull" the probability surface well into the central parts and down towards the eastern shores of the country.

Estimated variance

The se_{jack} and coefficient of variation associated with the best estimate for each grid cell are illustrated in Figs. 2a and 2b respectively.

The best estimate communicates a lot more information when interpreted in the context of its estimated variance. For example, cell A (Fig. 2b) has a probability of occurrence (best estimate) of 0.72 and a se_{jack} of 0.42. This large variation suggests that the best estimate is

(a)

Estimated standard error underlying predictions
- 0 - 0.125
- 0.126 - 0.25
- 0.251 - 0.375
- 0.376 - 0.5
- 0.51 - 0.7

600   0   600   1200 Kilometers

N

(b)

Cell A

Cell B

Estimated coefficient of variation underlying predictions
- 0 - 0.25
- 0.26 - 0.5
- 0.51 - 1
- 1.01 - 1.25
- 1.26 - 1.5
- 1.51 - 2.5
- 2.51 - 4.5
- Coefficient of variation undetermined for probability of occurrence = 0

600   0   600   1200 Kilometers

N

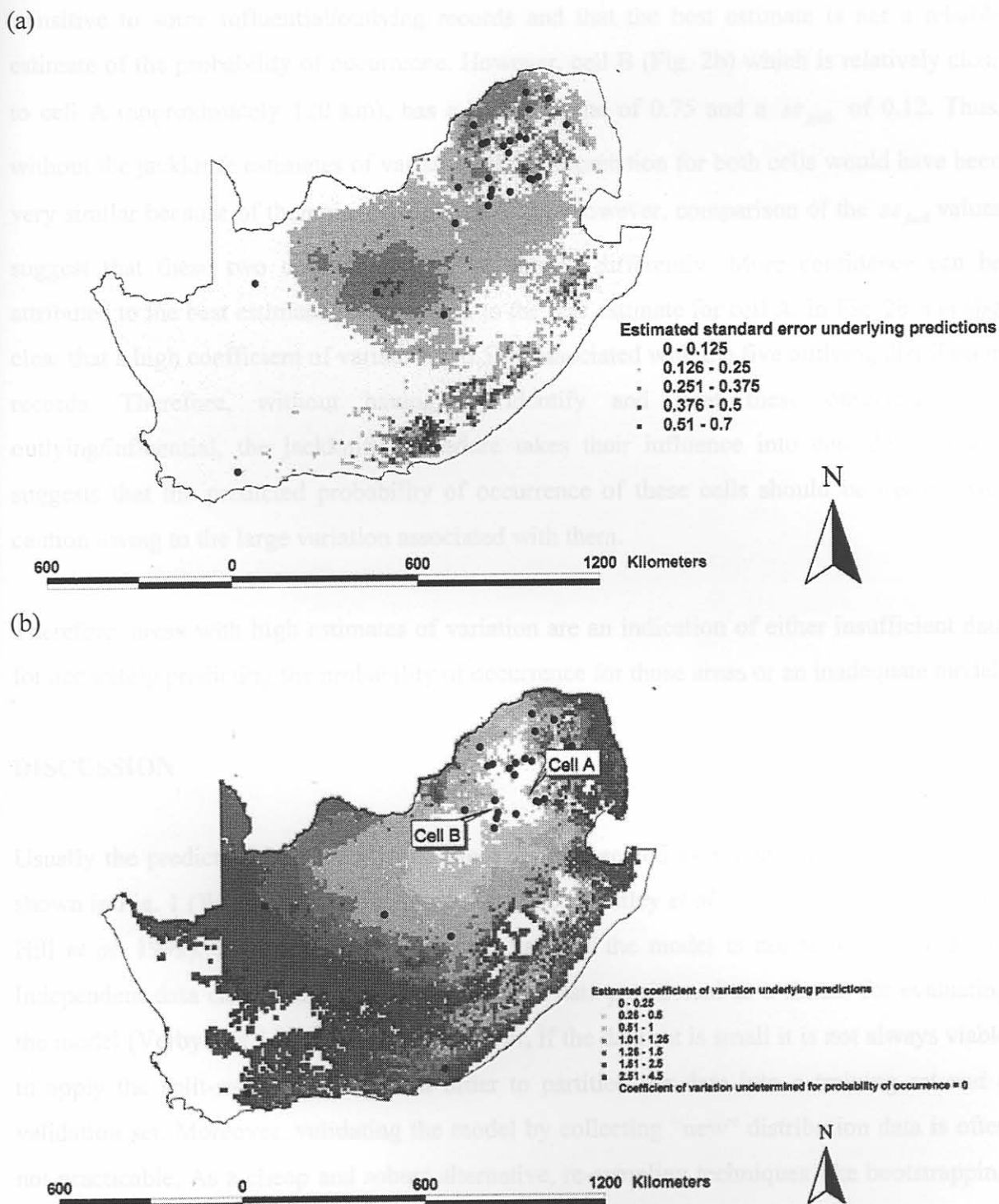Fig. 2: Known distribution records (•) of *Xylopa senior* and (a) the estimated standard error (se_{jack}), and (b) coefficient of variation underlying the predicted probability of occurrence surface for the species. The modelled probability of occurrence values for cell A and B are 0.72 and 0.75. The estimated standard error underlying each prediction is 0.42 and 0.12 for cells A and B respectively.

124

sensitive to some influential/outlying records and that the best estimate is not a reliable estimate of the probability of occurrence. However, cell B (Fig. 2b) which is relatively close to cell A (approximately 120 km), has a best estimate of 0.75 and a se_{jack} of 0.12. Thus, without the jackknife estimates of variation, the interpretation for both cells would have been very similar because of their similar best estimates. However, comparison of the se_{jack} values suggest that these two cells should be interpreted differently. More confidence can be attributed to the best estimate of cell B than to the best estimate for cell A. In Fig. 2b it is also clear that a high coefficient of variation (>0.5) is associated with the five outlying distribution records. Therefore, without having to identify and treat these observations as outlying/influential, the jackknife procedure takes their influence into consideration and suggests that the predicted probability of occurrence of these cells should be treated with caution owing to the large variation associated with them.

Therefore, areas with high estimates of variation are an indication of either insufficient data for accurately predicting the probability of occurrence for those areas or an inadequate model.

DISCUSSION

Usually the prediction of a bioclimatic model is represented as a single distribution map as shown in Fig. 1 (Walker 1990; Beerling *et al*. 1995; Huntley *et al*. 1995; Lindsay *et al*. 1998; Hill *et al*. 1999). Such maps are of limited value if the model is not tested or validated. Independent data can be collected or the original data partitioned as a means for evaluating the model (Verbyla & Litvaitis 1989). However, if the data set is small it is not always viable to apply the split-sample approach in order to partition the data into a training set and a validation set. Moreover, validating the model by collecting "new" distribution data is often not practicable. As a cheap and robust alternative, re-sampling techniques like bootstrapping and jackknifing can provide valuable insight into model performance (Verbyla & Litvaitis 1989; Osborne & Tigar 1992; Fielding & Bell 1997; Manel *et al*. 1999, Guisan & Zimmerman 2000). For example, when modelling techniques using presence/absence data instead of presence only data are used (e.g. logistic regression), real values can be compared to values predicted by the jackknife procedure, using the prediction from the model without the observation considered..The overall predictive ability of the model can then be evaluated by applying, for instance, the ROC-plot approach (Fielding & Bell 1997). Moreover,

although the value of re-sampling techniques for validating or testing distribution models has been recognised, they are still not routinely used. Here we have shown how re-sampling can be used effectively to provide a measure of confidence in model predictions, and how, in consequence, the interpretation of predicted distributions can be improved. Arguably, insights can be gained from a careful, expert-based scrutiny of the data; the main advantage of this jackknife procedure is that it provides quantitative support for such expert opinions, and where expert opnion is not readily available, it provides a method with which to assess to data of unknown quality.

Jackknifing may also to some extent be used to detect misclassifications of species, especially for species that are taxonomically closely related. For example, in the *Xylocopa senior* case study presented in this paper we suspected that five of the observations were misclassifications due to their geographical location. However, without having to treat these observations as misclassification or re-examining these records, the jackknife technique supported our suspicion by attributing high variation to the predicted probability of occurrence values at these sites and the surrounding areas (Fig. 2b).

It has to be borne in mind that there might be computational limitations when applying this jackknife approach to large datasets. There are more small data sets than larger ones, and the potential limitation of computing power on large data sets do not invalidate the usefulness of the approach on smaller data sets, where there are no computing limitations. The analyses in this paper was conducted on what can now be considered an entry level Pentium 3 workstation, and analysis on other data sets of up to 1000 records have been achieved within reasonable time limits

Arbitrary thresholds are sometimes placed on probability surfaces to identify "core" distribution areas of species (Walker 1990; Buckland & Elston 1993; Huntley *et al.* 1995). Fig. 3. illustrates all the grid cells with a probability of occurrence ≥ 0.5. We arbitrarily used this threshold value for illustration purposes, but methods for optimising this value can be found in the literature (e.g. Huntley *et al.* 1995, Zweig & Campbell 1993). The variation underlying predictions can be used as an additional condition to identify core distribution areas. In Fig. 3, all the filled cells have a probability of occurrence ≥ 0.5, but the dark grey
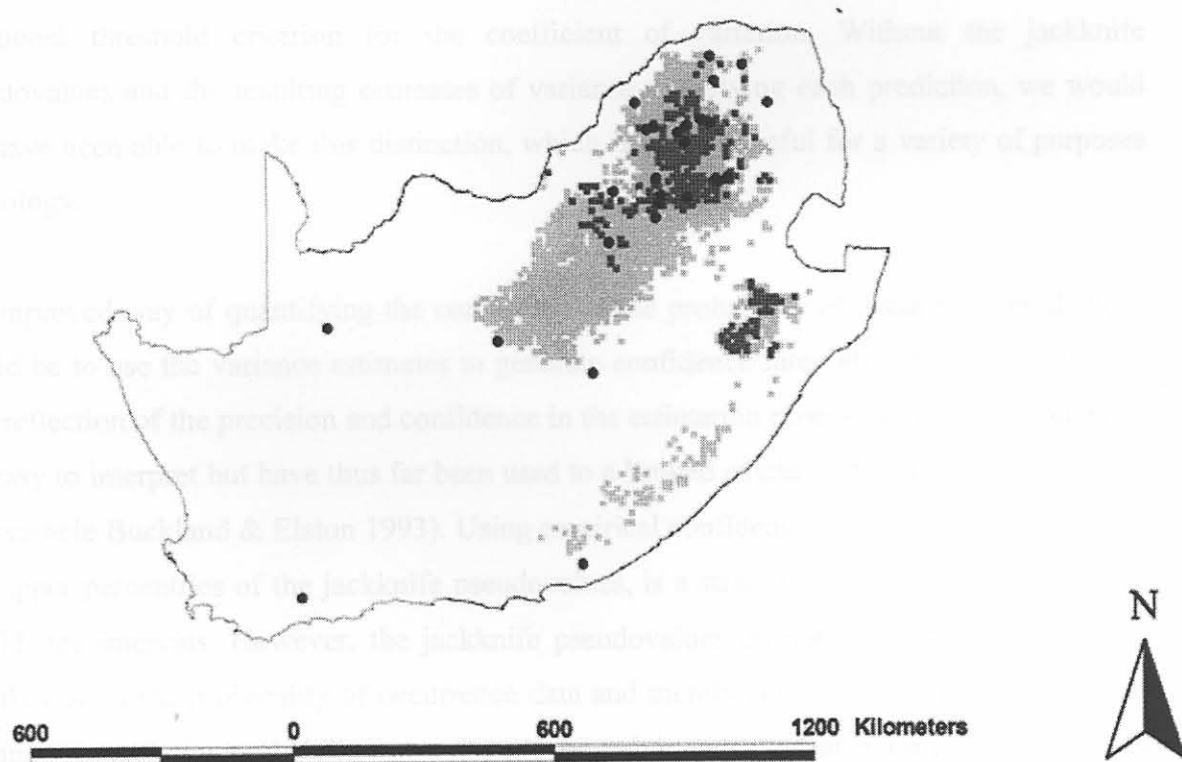
Fig. 3: Known distribution records (•) of *Xylocopa senior*. All the filled cells have predicted probability of occurrence values of ≥ 0.5. Dark grey cells satisfy the additional condition that the coefficient of variation underlying the predicted probability of occurrence is ≤ 0.15

cells satisfy the additional condition that the coefficient of variation is ≤ 0.15 (also arbitrary). Therefore, the dark grey cells are a filtered subset of the light grey cells, satisfying the additional threshold criterion for the coefficient of variation. Without the jackknife pseudovalues and the resulting estimates of variance underlying each prediction, we would not have been able to make this distinction, which is clearly useful for a variety of purposes in ecology.

An improved way of quantifying the confidence in the probability of occurrence predictions would be to use the variance estimates to generate confidence intervals. These intervals will be a reflection of the precision and confidence in the estimation process. Confidence intervals are easy to interpret but have thus far been used to a limited extent in distribution studies (see for example Buckland & Elston 1993). Using empirical confidence limits, based on the lower and upper percentiles of the jackknife pseudovalues, is a straightforward way of calculating confidence intervals. However, the jackknife pseudovalues do not simulate the underlying distribution of the probability of occurrence data and merely generates pseudovalues closely clumped around the probability as only one record is removed at a time. The empirical confidence intervals will therefore consistently underestimate the real interval length. The purpose of these jackknife pseudovalues is to estimate the variation (using equation 1) and they should not be used for calculating empirical confidence limits. Empirical confidence limits can only be used with bootstrapping, another re-sampling method, which simulates the underlying distribution of the statistic under consideration (this approach was followed by Buckland & Elston 1993). Since empirical confidence limits cannot be used for jackknifing, the normal approximation is sometimes used. Equation 2 gives the upper and lower confidence limits of a jackknife confidence interval if it can be assumed that $\hat{\theta}_{best\ estimate}$ is normally distributed:

$$\hat{\theta}_{best\ estimate} \pm \left\{ z_{\alpha/2} \times se_{jack} \right\} \qquad \text{(eqn 2)}$$

where: $\hat{\theta}_{best\ estimate}$ = best estimate;

$z_{\alpha/2}$ = $(1 - \alpha/2)$ th percentile of the standard normal distribution;

se_{jack} = standard error of the best estimate (from equation 1)

One of the assumptions of equation 2 is that the underlying distribution of the statistic under consideration is approximately normal (Efron & Tibshirani 1993). This condition is not true for a statistic that is bounded, like the probability of occurrence values generated in the

present study ($0 \leq \theta_{best\ estimate} \leq 1$). Where a statistic is bounded, transformations can be used to normalise the data. For example, Efron & Tibshirani (1993) used the Fisher transformation to normalise jackknife pseudovalues before confidence limits were calculated for correlation coefficients (a correlation coefficient, ρ, is bounded: $-1 \leq \rho \leq 1$). We tested a range of transformations, inter alia the probit (Cox 1970) and logit (Williamson & Gaston 1999) transformations that are suitable for probability values that are bounded between zero and one. After transforming the raw data, histograms for jackknife pseudovalues for each grid cell were drawn and a Kolmogorov-Smirnov normality test employed to test whether the data were adequately transformed to an approximate normal distribution. However, all of the transformations we employed failed to normalise the data adequately. Further work is needed in this regard since confidence intervals, as a function of the estimated variation, will be easier to present and interpret.

The predicted probabilities of occurrence and their variance estimates can have important implications for ecological surveys, conservation and management. The jackknife procedure presents a quantitative way to identify outliers in known distribution records. Such outliers might be real and truly peripheral to the core distribution, or they may indicate some error in data collection. Either way, further attention to these specific records is warranted.. The same argument can be applied to the planning of productive sampling sites, conservation hotspot identification (Meyers *et al.* 2000) and reserve selection (Margules & Pressey 2000). Therefore, without having to conduct extensive surveys, probabilities of occurrence together with their variance estimates can be used as a source of reference, as a decision-making tool in planning and for development proposals (Buckland & Elston 1993).

Another potential application of bioclimatic models is to predict possible distribution changes that may result from changing climates (Lindenmayer *et al.* 1991; Jeffree & Jeffree 1994; Huntley *et al.* 1995; Jeffree & Jeffree 1996; Sykes & Prentice 1995; Sykes *et al.* 1996; Hill *et al.* 1999). The interpretation of the predicted distributions under various climatic scenarios can be considerably improved by supplementing this information with indications of the variance underlying such predictions. Data sets with few known records may give rise to a probability of occurrence surface associated with large variance parameters, especially when the climate predictors for the different records are dissimilar.

Although the advantages of jackknifing is illustrated in this paper using a specific modelling approach, it is important to note that this re-sampling technique can readily be applied to most existing modelling techniques (e.g. GLM, GAM, CART, etc.). This makes jackknifing an important tool for estimating the variation underlying the predictions of various bioclimatic modelling techniques.

ACKNOWLEDGEMENTS

REFERENCES

Austin, M.P. & Meyers, J.A. (1996) Current approaches to modelling the environmental niche of eucalypts: implications for management of forest biodiversity. *Forest Ecol. Manag.*, 85, 95-106.

Austin M.P., Nicholls A.O. & Margules C.R. (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Monogr.*, 60(2), 161-177.

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

Brzeziecki B., Kienast F. & Wildi O. (1995) Modelling potential impacts of climate change on the spatial distribution of zonal forest communities in Switzerland. *J. Veg. Sci.*, 6, 257-268.

Buckland S.T. & Elston D.A. (1993) Empirical models for the spatial distribution of wildlife. *J. Appl. Ecol.*, 30, 478-495.

Carey P. & Ullyett J. (1993) Modelling the effects of climate change of species' distribution. *NERC News*, 24-25.

Caughley G., Short J., Grigg G.C. & Nix H. (1987) Kangaroos and climate: an analysis of distribution. *J. Anim. Ecol.*, 56, 751-761.

Chown S.L & Gaston K.J. (1999) Exploring links between physiology and ecology at macro-scales: the role of respiratory metabolism in insects. *Biol. Rev.*, 74, 87-102.

Cox D.R. (1970) *The analysis of binary data*. Chapman & Hall, New York.

Efron B & Tibshirani R.J. (1993) *An introduction to the Bootstrap*. Chapman and Hall, New York.

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

Erasmus, B.F.N., Van Jaarsveld, A.S., Chown, S.L., Kshatriya, M. & Wessels, K.J. (2002). Vulnerability of South African animal taxa to climate change. *Glob. Change Biol.*, 8, 679-693.

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

Guisan A. & Zimmerman N.E. (2000) Predictive habitat distribution models in ecology. *Ecol. Model.*, 135, 147-186.

Hill J.K., Thomas C.D. & Huntley B. (1999) Climate and habitat availability determine 20th century changes in a butterfly's range margin. *P. Roy. Soc. Lond. B Bio.*, 266, 1197-1206.

Huntley B., Berry P.M., Cramer W. & McDonald A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response surfaces. *J. Biogeogr.*, 22, 967-1001.

Jeffree C.E. & Jeffree E.P. (1996) Redistribution of the potential geographical ranges of Mistletoe and Colorado Beetle in Europe in response to the temperature component of climate. *Funct. Ecol.*, 10, 562-577.

Jeffree E.P. & Jeffree C.E. (1994) Temperature and the biogeographical distribution of species. *Funct. Ecol.*, 8, 640-650.

Johnson S.D. (1997) Pollination ecotypes of *Satyrium hallackii* (Orchidaceae) in South Africa. *Bot. J. Linn. Soc.*, 123, 225-235.

Koch S.O., Chown S.L., Davis A.L.V., Endrödy-Younga S. & van Jaarsveld A.S. (2000) Conservation strategies for poorly surveyed taxa: a dung beetle (Coleoptera, Scarabaeidae) case study from southern Africa. *J. Insect Conser.*, 4, 45-56.

Lindsay S.W., Parson L. & Thomas C.J. (1998) Mapping the ranges and relative abundance of the two principal African malaria vectors, *Anopheles gambiae sensu stricto* and *An. arabiensis*, using climate data. *P. Roy. Soc. Lond. B Bio.*, 265, 847-854.

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

Manel S., Dias J.M., Buckton S.T. & Ormerod S.J. (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *J. Appl. Ecol.*, 36, 734-747.

Margules C.R. & Pressey R.L. (2000). Systematic conservation planning. *Nature*, 405, 243-253.

Margules C.R. & Stein J.L. (1989) Patterns in the distribution of species and the selection of nature reserves: an example from *Eucalyptus* forests in South-eastern New South Wales. *Biol. Conserv.*, 50, 219-238.

Meyers N. , Mittermeier R.A., Mittermeier C.G., Fonseca G.A.B. & Kent J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853 – 858.

Osborne P.E. & Tigar B.J. (1992) Inte rpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *J. Appl. Ecol.*, 29, 55-62.

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

Shao J. & Tu D. (1995) *The Jackknife and Bootstrap*. Springer-Verlag, New York.

Sokal R.R. & Rohlf F.J. (1981) *Biometry. The Principles and Practice of Statistics in Biological Research.* 2nd edn. W.H. Freeman, San Francisco.

Sykes M.T. & Prentice I.C. (1995) Boreal forest futures: modelling the controls on tree species range limits and transient responses to climate change. *Water Air Soil Poll.*, 82, 415-428.

Sykes M.T., Prentice I.C. & Cramer W. (1996) A bioclimatic model for the potential distributions of north European tree species under present and future climates. *J. Biogeogr.*, 23, 203-233.

Verbyla D.L. & Litvaitis J.A. (1989) Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environ. Manage.*, 13, 783-787.

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

Williamson M. & Gaston, K.J. (1999) A simple transformation for sets of range sizes. *Ecography*, 22, 674-80.

Zweig, M.H. & Campbell, G. (1993) Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 39(4), 561-577.