DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals

G. CARPENTER, A.N. GILLISON[1]*, J. WINTER[2]

[1]*CSIRO Tropical Forest Research Centre, Atherton, Queensland, Australia*
[2]*P.O. Box 151, Ravenshoe, Queensland 4872, Australia*

This paper briefly reviews some limitations associated with the application of existing modelling procedures to conservation ecology, and describes a new procedure (DOMAIN) which avoids these problems. The procedure computes potential distributions based on a range-standardized, point-to-point similarity metric and provides a simple, robust method for modelling potential distributions of plant and animal species. DOMAIN offers advantages over similar methods in its ability to operate effectively using only presence records and a limited number of biophysical attributes. The use of a continuous similarity function gives DOMAIN increased flexibility as an heuristic tool, suitable for application in survey design, reserve selection and potential mapping of rare and common species. Potential distributions were computed for two Australian marsupial bettong species (*Aepyprymnus rufescens* 'Gray' and *Bettongia tropica* Wakefield) using DOMAIN and two alternative models. Of the three procedures, the DOMAIN model produced distribution patterns that were most consistent with the known ecology of the species, and most appropriate for survey design.

Keywords: spatial modelling; potential mapping; distribution mapping; Bettongs

Introduction

Understanding the distribution of key taxa and their response to environmental change is fundamental to effective management of the natural living resource. Because only a limited sample of the actual distribution of a taxon is usually achieved in practice, ecologists and conservation managers alike rely increasingly on spatial models to indicate the potential distribution range of taxa and the likely consequences of environmental impacts on selected species groups. Such distribution models are becoming important tools for survey design, reserve selection and management of rare species (Gillison and Carpenter, 1992).

Distribution models may be grouped into two broad classes: those which simulate detailed interactive processes between organisms and their environment using explicit performance parameters; and those which use pattern analysis techniques to exploit the correlation between environmental variables and the sampled distribution of target taxa.

Process models demand a detailed knowledge of organism response to a given range of environmental factors over time. For example, the CLIMEX model of Sutherst and

---

*To whom correspondence should be addressed at: P.O. Box 780, Atherton Qld 4883, Australia.

Maywald (1985), is capable of comparing the relative growth potential and persistence of populations of poikilothermic organisms at different times and places. It requires knowledge of animal response to given environmental extremes together with detailed climatic records. Ecologists tend to concentrate much effort into such models to elucidate spatial and temporal patterns of organism behaviour and system dynamics. See also Shugart *et al.* (1980), Webb *et al.* (1980) and Booth (1991). While providing important biological and ecological insights these models do not necessarily satisfy the immediate needs of conservation, where adequate process information is usually lacking.

The other broad class of models uses distribution records as surrogates for explicit organism performance parameters. Relatively modest data requirements allow this class of model to be applied to a wide variety of ecological problems requiring estimates of potential distribution (Box, 1981a, b; Booth, 1985; Busby, 1986; Nix, 1986; Booth and Jovanovic, 1988; Nix and Switzer, 1991; Walker and Cocks, 1991; Nix *et al.*, 1992).

In Australia, one very widely applied model in this class, BIOCLIM (Busby, 1986; Nix, 1986), uses a set of climate attributes for each site to derive a profile of suitable habitat based on limits observed for each attribute. Recent applications use 22 different
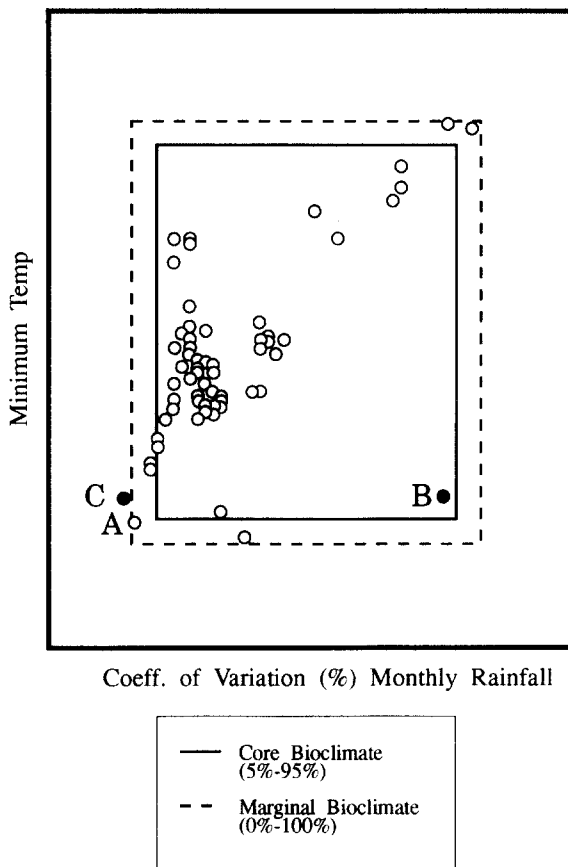

Figure 1. Boxcar environmental envelope.

climate attributes for each site. The wide acceptance and application of BIOCLIM underlines the need for a generic correlative modelling tool.

Some deficiencies have been noted in BIOCLIM; the range-based model treats each climatic axis independently, leading in some cases to ecologically unsound predictions (Booth, 1990). Potential problems can be illustrated using a sample of training points plotted against two climatic axes (Fig. 1). Point A, a known record site, is excluded from the 'core bioclimate'. As the climatic envelope is constricted toward the midpoint by selecting narrower percentile ranges, an increasing number of record sites is excluded from the predicted distribution. Point B, which is comparatively dissimilar to all known record sites is included in the 'core bioclimate'. In ecological terms this could mean records of an organism in both hot-wet and cool-dry environments are interpreted by BIOCLIM as implying suitability of cool-wet environments and hot-dry environments. It is also significant that point C, which is very similar to known record site A, is excluded from the environmental envelope. This would result in a site which is only marginally wetter than the wettest record site being completely excluded from the predicted distribution map.

Walker and Cocks (1992) noted limitations in BIOCLIM and proposed an alternative procedure (HABITAT) which uses the convex hull of record sites in environmental space to more tightly define the limits of suitable bioclimate. HABITAT employs Classification and Regression Trees (CART; Brieman *et al*, 1984) and species absence records to aid in the selection of key climatic parameters and to dissect the environmental envelope into sub-volumes of varying classification confidence.

The convex hull model redresses some of the classification problems identified in BIOCLIM, correctly encompassing known record site A, and rejecting point B (Fig. 2). As with BIOCLIM, the HABITAT envelope excludes point C, despite its close proximity to record site A.

Materials and methods

Three correlative distribution models

Boxcar The BIOCLIM procedure identifies locations where all climatic indices fall within the extreme values determined for a set of observation records. Multiple levels of classification are achieved by identifying locations with climatic values contained within fractional parts of the total range. For example, in a two-level classification using 0–100 and 5–95 percentile limits, a 'core bioclimate' would contain those locations where all climatic indices fall between the 5- and 95-percentiles and a 'marginal bioclimate' would contain those locations where one or more of the climatic indices fall outside the 5- and 95-percentiles, but with all indices remaining within the outer limits.

Thus BIOCLIM defines the environmental envelop for a target taxon as a rectilinear volume in a Euclidean space, as illustrated in Fig. 1. As range limits are decreased, the volume constricts toward the midpoint of the record sites in climate space. The simplicity of the BIOCLIM envelope permits it to be concisely described as a table of range values. This range-based classification method, also used in other modelling systems (Box, 1981a; Booth, 1991), is analogous to the 'parallelepiped' or 'boxcar' image classification algorithm, widely applied in multi-spectral remote sensing applications.

The boxcar model tested here uses the BIOCLIM algorithm, differing only in the number of available data layers. Whereas BIOCLIM is typically applied using more than
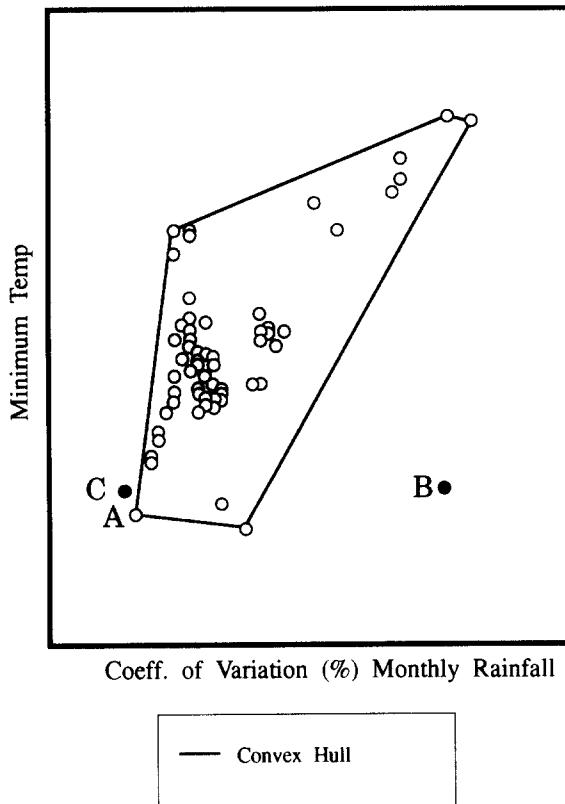
Figure 2. Convex hull environmental envelope.

20 climate variables, our boxcar model is restricted to a set of 5 climate attributes available for our study area.

Convex hull   Instead of a rectilinear volume in environmenal space, HABITAT uses the convex hull of the training sites to more tightly constrain the environmental envelope (Fig. 2). HABITAT applies CART to discard those parameters least capable of discriminating between presence and absence sites, and again to subdivide the polytope volume into regions of varying classification confidence based on proportions of presence and absence sites.

The convex hull environmental envelope used by HABITAT offers one solution to the inter-dimensional independence problem identified in BIOCLIM. To test the utility of this type of environmental envelope in applications lacking absence data, we created a model which performs binary classification by determining if candidate sites are included in the convex hull of the record sites. Thus our model resembles HABITAT but without the two CART stages.

DOMAIN   The DOMAIN procedure uses a point-to-point similarity metric to assign a classification value to a candidate site based on the proximity in environmental space of
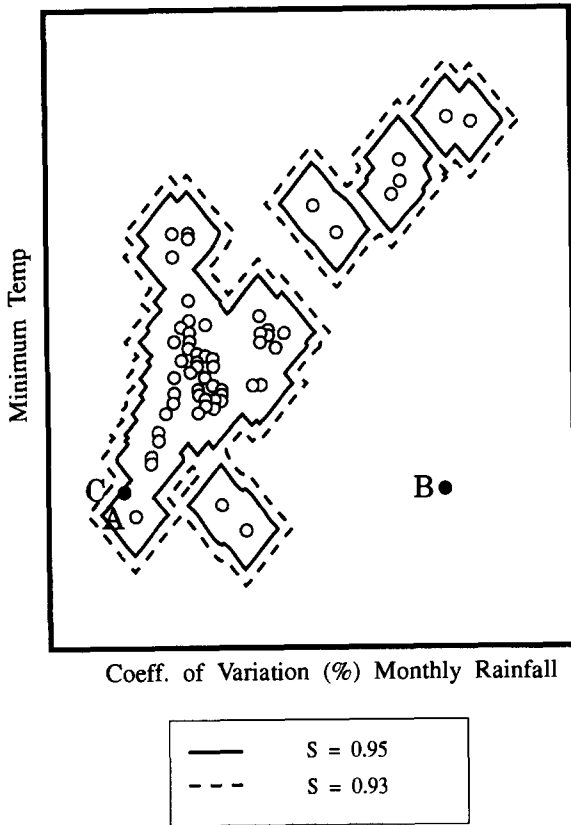
Figure 3. DOMAIN environmental envelope.

the most similar record site (Fig. 3). This is similar to the method proposed by Booth (1990) for SITESIM, a programme for evaluating site suitability for plantation trials. Belbin (1992) employed a somewhat similar point-to-point algorithm to measure similarity between two community species profiles.

The Gower metric (Gower, 1971) provides a suitable means of quantifying similarity between two sites. The distance of d between two points A and B in a Euclidean p-dimensional space is defined as:

$$(1) \qquad d_{AB} = \frac{1}{p} \sum_{k=1}^{p} \left(\frac{|A_k - B_k|}{\text{range}k} \right)$$

The Gower metric uses range standardization to equalize the contribution from each climatic attribute. This method of standardization is preferred over variance standardization in this application because it is less susceptible to bias arising from dense clusters of sample points.

We define the complementary similarity measure R_{AB}:

$$(2) \qquad R_{AB} = 1 - d_{AB}$$

R is constrained between 0 and 1 for points within the ranges used in Equation 1, but may yield negative values for points outside this range. We define S_A, the maximum similarity between candidate point A and the set of known record sites T_m as:

$$(3) \qquad\qquad\qquad S_A = \max_{j=1}^{m} R_{T_jA}$$

By evaluating S for all grid points in a target area, a matrix of continuously varying similarity values is generated which can be displayed as a grey-scale, contour or thematic map. As with all models discussed here, the values generated are not probability estimates, but degrees of classification confidence. DOMAIN defines no discrete boundary for the climate envelope. All candidate points are assigned similarity values and user-defined thresholds or contour intervals determine the actual ranges mapped. For example, a threshold of $S \geq 0.90$ would select points with an average variation in climate values from a known occurrence site of no more than 10% of the range.

Threshold values for a particular application may be based on expert knowledge, or a number of subjective thresholds may be used to reveal relative trends. Absence records, when available, can be used to establish thresholds more objectively by determining the value which minimizes the classification error.

A comparative study of Bettong distribution using Boxcar, Convex hull and DOMAIN

Site records for two north Queensland Bettong species were chosen for analysis. Because the two species are essentially allopatric (i.e. they occupy mutually exclusive domains) it was felt the extent to which each model predicted species overlap would serve as a relative measure of ecological utility. The set of training points consisted of 20 presence records for the northern Bettong (*Bettongia tropica*) and 71 presence records for the Rufous Bettong (*Aepyprymnus rufescens*), the latter representing the eastern margin of the distribution of this species. Climate data were available with a grid spacing of 12 seconds (approximately 360 m) providing 580 184 terrestrial candidate sites with the following five climate variables:

(i)   daily maximum air temperature averaged over hottest month,
(ii) daily minimum air temperature averaged over coldest month,
(iii) coefficient of variation of average monthly rainfall,
(iv) total rainfall of three driest successive months; and
(v)   average annual rainfall.

Results

Boxcar

The resulting maps of the core bioclimate regions from the boxcar procedure (Fig. 4) encompass comparatively large areas, and include regions regarded by J. Winter as being unlikely to contain the target taxa. Of the sites included in the core bioclimate for *B. tropica*, 95% were also included in the *A. rufescens* core bioclimate. This large spatial overlap (Fig. 8) is not supported by current understanding of the ecology of these primarily allopatric species.
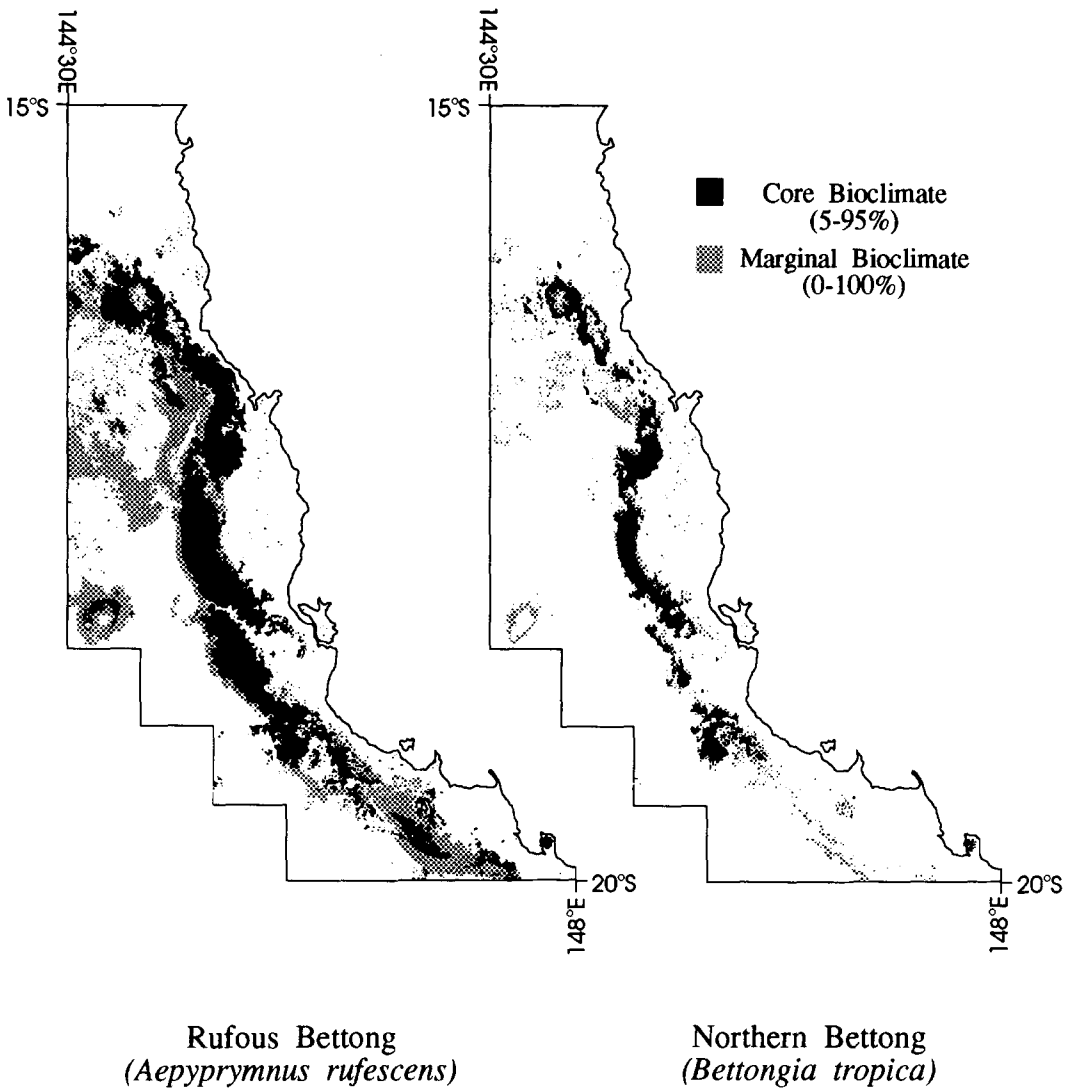
Rufous Bettong
(Aepyprymnus rufescens)

Northern Bettong
(Bettongia tropica)

Figure 4. Boxcar projected distribution.

It appears that this distribution pattern is a product of the complete inter-dimensional independence of the method when applied to highly correlated climate data. Similar problems are encountered in applying parallelepiped classification to covariant image data (Lillesand and Keifer, 1987).

Convex hull

The conservative spatial distributions predicted from the convex hull model (Fig. 5) are of limited use to ecologists, reflecting little more than the original collection sites. This severely restricted projection can be attributed to the very rapid reduction in the hyper-
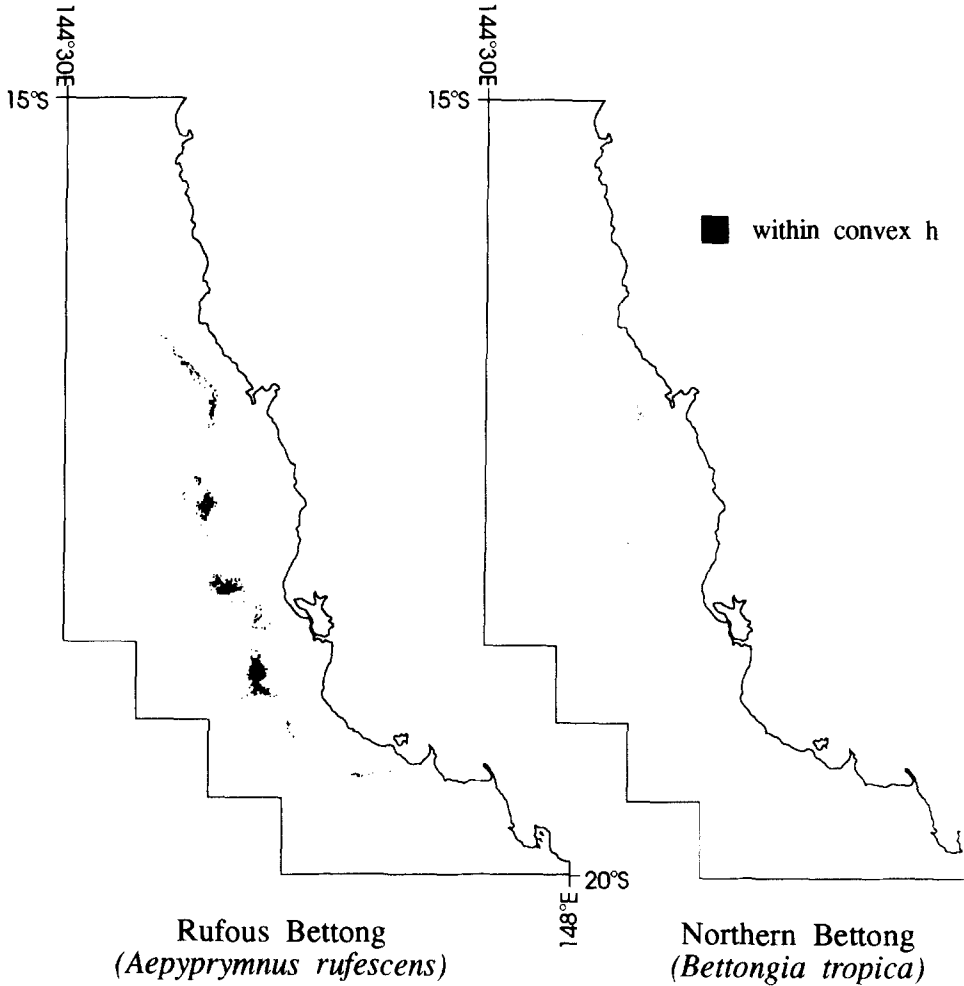
Figure 5. Convex hull projected distribution.

volume delineated by the convex hull as dimensionality increases (Jancey, 1980). Using five dimensions, the convex hull encompasses the record sites so tightly that it virtually excludes any variation in conditions from the mapped distribution.

DOMAIN

The spatial predictions produced by DOMAIN are shown in Fig. 6. To gain some insight into likely distributional margins between the two species, map thresholds were selected to maximize the predicted distribution area while minimizing the distributional overlap in accordance with the known allopatry of the two species. This was achieved by generating a frequency distribution of S values (Fig. 7) and selecting the lowest threshold ($S \geqslant 0.95$) with no significant overlap (Fig. 8). A second threshold ($S \geqslant 0.93$) was arbitrarily selected to illustrate additional similarity trends.
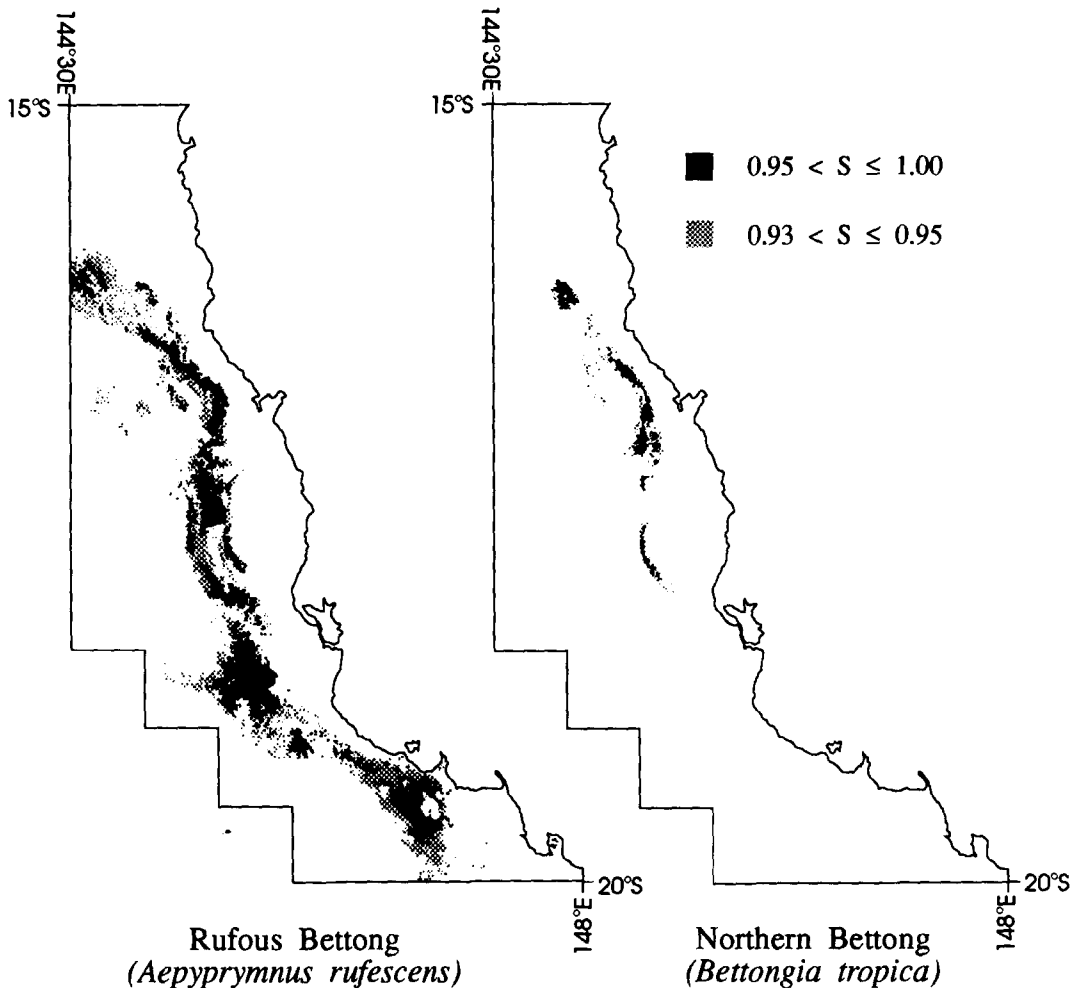
Figure 6. DOMAIN projected distribution.

Discussion

The data set used to compare the three models relies on a limited number of climate variables which, although demonstrably important for plant and animal distribution (Nix and Switzer, 1991), may not be the most relevant for modelling Bettong performance. There can be no adequate test of the mapped distributions other than by subsequent survey of the predicted domain. Nevertheless the large degree of variation in the results offers insight into the performance characteristics of these three models.

Because the threshold chosen for DOMAIN was purposively derived to generate allopatric distributions, the resulting allopatry does not itself support the DOMAIN approach. Subsequent field sampling of the predicted distribution would provide feedback to refine threshold settings for these species and improve the accuracy of subsequent predictions.
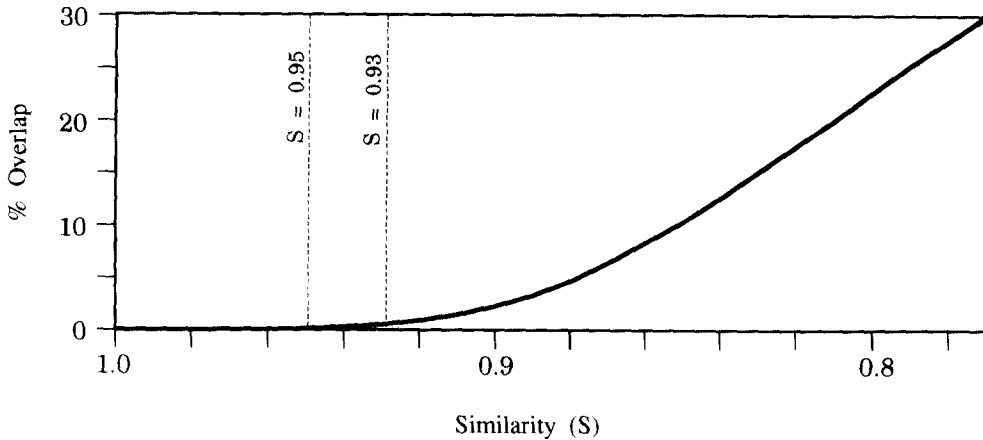
Figure 7. Range overlap between species.

Of the three models tested, DOMAIN generated the distribution most appropriate for subsequent survey design and ground-truthing, being consistent with the known ecology of the species (unlike boxcar) and sufficiently liberal to suggest additional areas for sampling (unlike the convex hull).

DOMAIN calculates a continuous similarity function for all candidate sites. Although not conventionally expressed as such, the boxcar model can also be formulated as a continuous function. Whereas the DOMAIN function measures the environmental similarity to the most similar training site, the boxcar function measures the distance to the midpoint of the training sites in climate space. The difference in practice between these functions can be illustrated in an area in the south-west corner of the map corresponding to the St. Ronans 1:100 000 map sheet, where the boxcar model identified a conspicuous core bioclimate isolate (Fig. 4). Core bioclimate mappings reflect sites with climate attributes varying on average by no more than 45% of range from the midpoint of the training sites, and by definition some record sites will necessarily fall outside the core bioclimate. Thus the St. Ronans isolate is actually given a higher classification confidence than some known record sites. DOMAIN gives the isolate considerably less weight. A contour map of DOMAIN *S* values (Fig. 9) reveals a local maximum in this area with similarity values above 0.90, and below 0.92. This implies climatic conditions in this region vary between 10% and 8% of range from a known record site, compared with a variation of less than 5% for the primary distribution mapped in Fig. 6.

In some cases sample data may contain significant errors due to factors such as miscoded site locations, sketchy anecdotal records, wrong identifications from poor or non-vouchered material or subsequent nomenclatural changes in taxonomic revisions. The impact of spurious outliers arising from such problems can be reduced by directing DOMAIN to use the mean of a number of largest similarity values rather than the maximum similarity value.

DOMAIN lends itself to a range of heuristic applications beyond that of predicting potential distributions. For example, inadequacies in a proposed survey design may be exposed by using the set of proposed survey sites as a DOMAIN training set, and
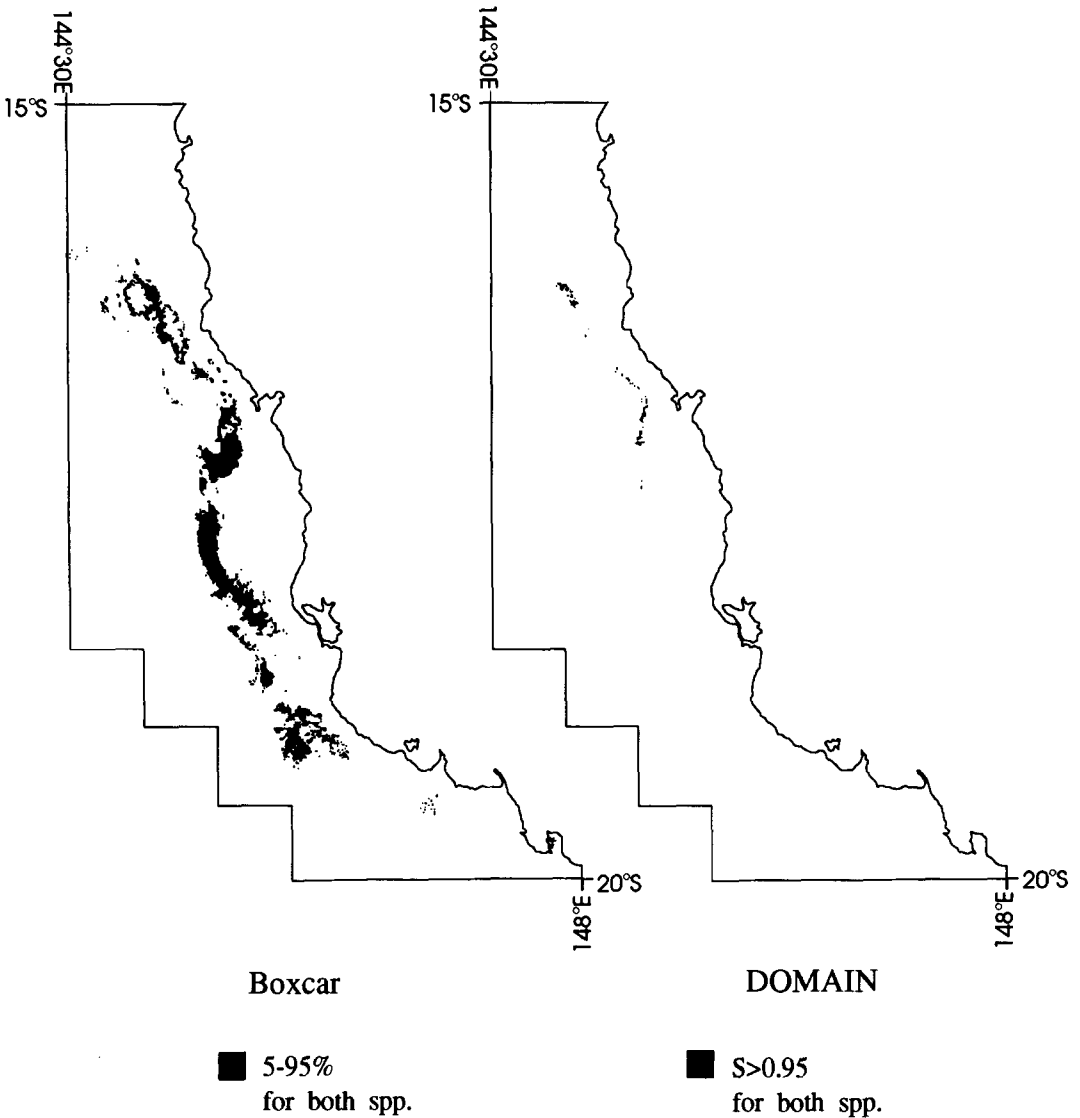
Figure 8. Interspecies overlap for Boxcar and DOMAIN.

mapping regions with climatic conditions that differ significantly from those covered by the sample design. Similarly, reserve representativeness can be tested by using a set of points enclosed within a proposed reserve as training sites and mapping non-reserve regions with the greatest degree of dissimilarity.

Conclusions

The use of a point-to-point similarity metric offers significant advantages over spatial models that rely on rectilinear or convex hull environmental envelopes (Table 1). DOMAIN provides a useful complementary alternative for general potential mapping,
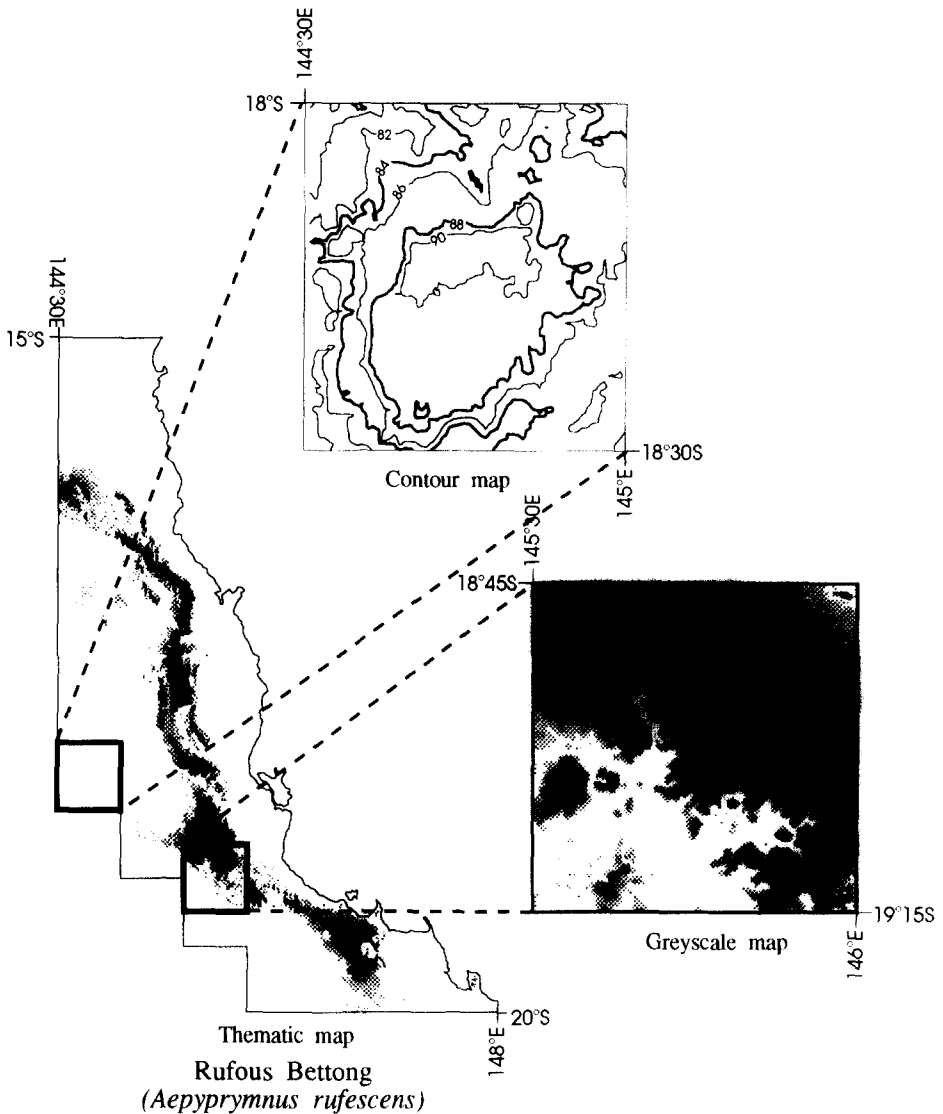
Figure 9. DOMAIN distribution: thematic, contour and greyscale maps.

and is particularly well suited to applications where available site location records or environmental data are limited.

We have recently acquired presence and absence data from an intensive survey of plant and animal distributions along a series of well documented physical environmental gradients in the wet tropics of north Queensland. These will help develop the DOMAIN algorithm and will form the substance of a second paper to focus on applications in conservation management.

Table 1. Comparison of tested models

Model	Classification method	Features
Boxcar	multilevel rectilinear envelope	• dimensions treated independently • performs poorly on covariant data • includes some dissimilar sites • excludes some similar sites • easily implemented • simply described envelope
Convex hull	binary convex hull envelope	• tightly constrained envelope • excludes many similar sites • difficult to implement • computationally expensive
DOMAIN	continuous point-to-point similarity metric	• variable sensitivity • performs well with limited site data • gives similarity value to all sites • easily implemented

Acknowledgements

References

Belbin, L. (1992) Comparing two sets of community data: A method for testing reserve adequacy. *Aust. J. Ecol.* **27**, 255–62.

Booth, T.H. (1985) A new method for assisting species selection. *Commonw. For. Rev.* **64**, 241–50.

Booth, T.H. (1990) A climatic analysis method for expert systems assisting three species introductions. *Agroforestry Syst.* **10**, 33–45.

Booth, T.H. (1991) Plant growth and site selection: where will it grow? how will it grow? *Agric. Syst. Inform. Technol. Newslett.* **3.3**, 30–1.

Booth, T.H. and Jovanovic, T. (1988) Assaying natural climatic variability in some Australian species with fuelwood and agroforestry potential. *Commonw. For. Review* **67**, 27–34.

Box, E.O. (1981a) *Macroclimate and Plant Forms: An Introduction to Predictive Modelling in Phytogeography*. The Hague: Dr W. Junk.

Box, E.O. (1981b) Predicting physiognomic vegetation types with climate variables. *Vegetatio* **45**, 127–39.

Brieman, L., Freidman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees*. Belmont: Wadsworth.

Busby, J.R. (1986) A biogeographical analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. *Aust. J. Ecol.* **11**, 1–7.

Gillison, A.N. and Carpenter, G. (1992) A regional sampling strategy for natural resource inventories within the Wet Tropics World Heritage Area. Unpublished report to Wet Tropics Management Agency, Cairns, Queensland.

Gower, J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics* **27**. 857–71.

Jancey, R.C. (1980) A hyperspatial model for complex group structure. *Taxon* **26**, 409–11.

Lillesand, T.M. and Keifer, R.W. (1987) *Remote Sensing and Image Interpretation*, 2nd edition. New York: John Wiley & Sons.

Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. In *Atlas of Elapid Snakes of Australia* (R. Longmore, ed.) Canberra: AGPS.

Nix, H.A., Stein, J.A. and Stein, J.L (1992) *Developing an Environmental Geographic Information System for Tasmania*. Canberra: Centre for Resource and Environmental Studies, Australian National University.

Nix, H.A. and Switzer, M.A., eds (1991) *Rainforest Animals, Atlas of Vertebrates Endemic to Australia's Wet Tropics* in Kowari No. 1. Canberra: Australian National Parks and Wildlife Service.

Shugart, H.H.,, Mortlock, A.T., Hopkins, M.S. and Burgess, I.P. (1980) A computer simulation model of ecological succession in Australian subtropical rainforest. Oak Ridge National Laboratory, ORNL/TM 7029.

Sutherst, R.W. and Maywald, G.F. (1985) A computerised system for matching climates in ecology. *Agric. Ecosyst. Environ.* **13**, 281–99

Walker, P.A. and Cocks, K.D. (1991) HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecol. Biogeog. Lett.* **1**, 108–18.

Webb, D.B., Wood, P.J. and Smith, J.P. (1980) A guide to species selection for tropical and subtropical plantations. *Tropical Forestry Paper No 15*. Oxford: Commonw. For. Inst.