Department of Conservation and Land Management

Comprehensive Regional Assessment for a
Regional Forest Agreement for Western Australia

Distribution mapping and key ecological attributes of plant species in the south-west forest region

Final Report

by P. Gioia, J.P. Pigott, R.J.Cranfield & A.R. Chapman

May, 1998

Table of Contents

1 Introduction

This project represents part of the effort to assess and quantify aspects of biodiversity within the study area of the 1997 Comprehensive Regional Assessment (CRA) for the Regional Forest Agreement (RFA) process in Western Australia.

Biological diversity is usually considered at three levels:

- "Genetic diversity" refers to the variety of genetic information contained in all individual plants, animals and microorganisms.
- "Species diversity" refers to the variety of living species.
- "Ecosystem diversity" refers to the variety of habitats, biotic communities and ecological processes.

The Western Australian and Commonwealth Governments agreed that the South West Forest Region biodiversity assessment be undertaken at the species and ecosystem levels and include reviews of the inputs of the main disturbances to such biodiversity in the region. Because information about genetic variation within species is very limited and costly to obtain, genetic diversity was not assessed, although it is recognised that it does overlap with species and ecosystem diversity and these are addressed by the National Forest Reserve Criteria (JANIS 1997).

The biodiversity assessment has therefore been based on an analysis of information about forest ecosystems and communities, flora and fauna species and their habitats, and the inputs of disturbance on these in the region.

The CALM Herbarium was commissioned to undertake a number of projects specifically to study species distribution and reveal patterns of richness. (The development and analysis of ecosystem diversity was performed as part of other CRA projects.) The results of these projects would then provide the basis for "informing layers" to the integration and options development phase of the RFA.

During the initial planning phases of the assessment, three projects were devised for assessing flora and fauna diversity: the project reported on here, focussing on plant distribution and richness, and two related projects,

> *Detecting areas with high levels of faunal species richness* and
> *Distribution of species of special interest.*

Each of these projects was to be supervised by CALM Herbarium staff in collaboration with the Western Australian Museum. However, once the assessment was under way, a number of circumstances prevailed, resulting in a review and subsequent reduction of the scope and custodianship of each project.

From early September 1997, all plant-related analysis continued under the mantle of the CALM Herbarium. All fauna-related activities (including datasets such as that for scorpions and mammal species) were allocated to the WA Museum for analysis. Mapping of fauna species, using published distributions, for the above projects is documented in Abbott (1998).

This report documents only those aspects of each of the above projects that were undertaken specifically at the CALM Herbarium. Project objectives were to:

- validate existing Herbarium site records and collate with site records from vegetation mapping projects;
- undertake distribution mapping of agreed JANIS target species and species with reliable records;
- detect "hotspots" of plant species richness in WA forests in the context of the Southwest Land Division;
- determine key ecological attributes of species present in the southwest forest region.

The assessment of the South West Forest Region flora species involved the compilation and validation of distribution and ecological data from a range of sources, the development of the WABiota database and the analysis and display of various themes.

It was known in advance that the distribution of plant records within the RFA area would be subject to a range of biases. Some of these biases were characteristic of opportunistic collections, reflecting proximity to roads or favoured collecting sites. Other anticipated biases were the generally low level of species data available for the RFA area and the low collection rate of ubiquitous species. Statements on species diversity within the RFA area, relying solely on raw data, would reflect these biases with potentially misleading results.

The approach taken in this project was to model and predict the distribution of species based on existing records, compensating, to some extent, for the biases mentioned above. These predicted distributions then served as a basis for indicating areas of high species richness.

A detailed description of the methods involved in compiling plant distribution data, generating predicted distribution models and developing a species richness index for the RFA area is provided, together with an assessment of that index and its correct usage within the RFA process.

2 Methods

2.1 Data Review

The primary aim of a data review is to ascertain the adequacy of existing site-based biological data for determining the distribution of flora and fauna species, and relating this to their habitat requirements. Outcomes from the review can also be used to identify data gaps and priority areas for additional survey work. The Data Audit Methodology (DAM) software, developed and supplied by Environment Australia (Bennet, Bugg & Barratt, 1997), was designed to assist in the data review process by identifying areas for potential survey.

A review of the candidate datasets for the RFA process was attempted using the DAM software to identify possible gaps. For small subsets of data, the software initially provided promising results. However, the software proved unstable when using all available data across the entire study area. In part, instability was the result of certain constraints hardcoded into the software, such as the maximum number of species that could be analysed at one time. Subsequent versions of the software will no doubt remedy such trivial limitations.

2.2 Data Acquisition

Data were obtained from a variety of existing sources internal and external to CALM (see Section 3). Because species data from parts of the RFA area were known to be significantly under-represented, three major surveys in the RFA area were undertaken. These were a survey of selected sites for ecological attributes of vascular flora (see Ecological Attribute Field Survey below), a survey of vegetation complexes, and a survey of threatened taxa populations.

2.2.1 Ecological Attribute Field Survey

2.2.1.1 Project Objectives

- To determine key ecological attributes of species present in the south-west forest region of Western Australia.
- To augment existing information, collect, record and voucher material from selected forest sites for lodgement at the CALM Herbarium and to prepare data for inclusion into the CALM Herbarium Specimen Database.

2.2.1.2 Sampling Strategy

Field collecting aimed to sample every taxon of vascular plant from each of 150 sites. Vouchers of each species were collected together with information on their phenology, especially age since last fire, percentage of individuals flowering, presence of seed capsules and mode of perennation. Fire periods were to be confirmed from CALM records and in the field by observation on the morphology of selected species eg *Banksia grandis*, a common species widespread in the area.

2.2.1.3 Sampling and Recording of Field Data

Each site was selected using a series of maps prepared by CALM's Information Management Branch, depicting fire history for the past 10 years with surface contours to show uplands, lowlands and rainfall isohyets. The maps covered all three forest regions, eg Northern, Central and Southern forests, allowing up to 150 sites to be selected so that they covered a range of fire history, elevation and rainfall. An identification number was allocated to each site. Site locations were digitised onto CALM's 1:50,000 map series, requiring minor adjustments in some situations.

A recording sheet for use by field staff to capture site and biological information was developed in consultation with database operators and GIS staff. A specialised collecting book to record individual species data was prepared along with an information booklet to ensure uniformity between field teams. An ancillary database, WASite, was created and linked to WAHerb to manage detailed site data.

Initially, one team of two persons was employed to conduct the field sampling. This was later increased to 4 teams of 2 people. All field staff were trained to correctly record data and use global positioning system (GPS) units to accurately geocode the proposed sites. Each team was allocated a four wheel drive vehicle to gain access to the site area but since several sites were in disease risk areas, teams used tracks and located sites on foot, thus observing strict hygiene protocols.

At each site a circular plot of 100 metre radius (3.14 ha) was marked out using sighting poles and flagging tape. Site information data (soils, vegetation type, fire history, etc.) were gathered at the plot centre, with minor corrections and additions added after the site was completed. Samples were collected by firstly traversing the perimeter and then randomly sampling through the plot. Specimen vouchers were collected either as flowering or non flowering samples, tagged and pressed in preparation for identification. Percentage flowering of each species within and around the plot were recorded. Cryptogams (mosses, liverworts and lichens) found in the plot were also vouchered.

2.2.1.4 Identification and Databasing of Samples (Vouchers)

A team of three was employed to carry out the identification of the field vouchers of vascular plants. All material collected was identified either to genus or to species, with many identified to subspecies or variety levels. All identified vouchers were prepared for mounting and lodged with the State Collection at the CALM Herbarium. Vouchers were databased using standard protocols and assigned a unique, identifying bar code.

From the 118 sites a total of 13,410 vouchers were collected and, of these, 11,009 were databased for incorporation into WABiota (see Section 2.3). The remaining 2401 vouchers were cryptogams set aside for identification at a later date.

Table 1. Ecological Attributes Survey Site Information[1]

Northern Forest		Central Forest		Southern Forest	
Site	No. Taxa Collected All (Vascular)	Site	No. Taxa Collected All (Vascular)	Site	No. Taxa Collected All (Vascular)
1	121 (120)	51	107 (96)	101	68 (62)
2	118 (115)	52	107 (81)	102	
3	111 (103)	53	101 (82)	103	147 (123)
4	116 (108)	54	107 (93)	104	
5	129 (123)	55	97 (96)	105	65 (49)
6		56	107 (90)	106	116 (77)
7	142 (136)	57	121 (106)	107	
8	72 (68)	58	127 (112)	108	
9	71 (76)	59		109	
10	119 (117)	60	76 (72)	110	128 (94)
11	100 (97)	61	150 (134)	111	
12	184 (178)	62	154 (137)	112	96 (83)
13	178 (167)	63	95 (83)	113	72 (67)
14	142 (143)	64	84 (77)	114	
15	112 (111)	65	123 (108)	115	
16	107 (102)	66	84 (75)	116	106 (89)
17	98 (92)	67		117	
18	102 (92)	68	93 (81)	118	
19	109 (98)	69	164 (152)	119	
20	83 (80)	70	74 (67)	120	
21	130 (121)	71	87 (69)	121	
22	135 (100)	72	148 (132)	122	74 (53)
23	86 (82)	73	87 (75)	123	63 (58)
24	99 (87)	74	114 (95)	124	110 (105)
25	100 (91)	75		125	54 (44)
26	114 (110)	76	115 (90)	126	60 (58)
27	121 (120)	77	124 (104)	127	70 (60)
28	144 (126)	78	178 (151)	128	91 (80)
29	123 (111)	79		129	99 (95)
30		80	111 (92)	130	106 (103)
31	113 (97)	81		131	95 (90)
32	138 (132)	82	132 (104)	132	
33	120 (104)	83		133	112 (110)
34	121 (116)	84	150 (128)	134	94 (89)
35	160 (151)	85	149 (120)	135	67 (61)
36	106 (105)	86	112 (93)	136	107 (100)
37	112 (106)	87	110 (83)	137	74 (56)
38	106 (101)	88	172 (156)	138	
39	104 (96)	89		139	
40	94 (82)	90	173 (158)	140	
41		91		141	94 (93)
42	94 (79)	92	136 (119)	142	105 (100)
43	94 (85)	93	138 (127)	143	120 (115)
44	87 (61)	94	99 (86)	144	117 (105)
45	78 (75)	95	134 (109)	145	94 (89)
46	82 (76)	96	77 (59)	146	98 (96)
47		97	88 (60)	147	86 (71)
48	82 (70)	98		148	
49	78 (68)	99	134 (122)	149	101 (97)
50	75 (68)	100	105 (95)	150	

[1] Blank entries indicate sites which were not visited

2.2.2 Species Attribution

A database of species attributes was compiled from three main sources. Firstly, lists of taxa declared rare or marked as poorly known were obtained from CALM Wildlife Branch. Secondly, the Manjimup Research Centre had maintained a database of species response to disturbance, including fire and disease. There were a number of contributors to this database, including Neil Burrows (SID) and Grant Wardell-Johnson (formerly CALM). And, thirdly, Roger Hearn (Manjimup Region) compiled a list of attributes reflecting JANIS target groups including endemic taxa, relictual taxa and taxa at their range ends or having disjunct populations.

2.2.3 Other surveys

Voucher specimens for about 4000 vascular plant species collected for the *Vegetation Complex Mapping Project* (Mattiske Consulting Pty Ltd, 1997a) were lodged at the Herbarium. Specimen data were incorporated into WABiota. Additional vascular plant species data (un-vouchered) were provided by a number of mining companies and other agencies (see Table 4 below) for inclusion.

2.3 Development of WABiota

WABiota is a warehouse of flora data obtained from a variety of sources internal and external to CALM. It was designed to contain not only individual sight and specimen flora and fauna records but also taxonomic information on names and a variety of species-level ecological attributes as described in the previous section on Data Acquisition.

WABiota was constructed in Arc/INFO 7.1.1 on a Sun Ultra 2 running Solaris 2.5.1. The coordinate data were stored in standard Arc coverages and grids and textual data stored and managed within the INFO environment.

The database was supported by a number of scripts that enabled the updating of any given portion of the database from a recognised source such as WAHerb or the Declared Rare Flora (DRF) Database. The nomenclature for the majority of taxa was taken from WACensus, the authoritative database of plant names relevant to WA, as determined by the CALM Herbarium, the data custodian.

2.3.1 Validation of WABiota Records.

A method for prioritising the validation of geocodes was developed using climatic attributes and BIOCLIM (McMahon *et al.* 1995) to identify potential outliers as targets for validation. This was needed because of the large number of points within WABiota. A method was required that could automatically determine which points might require validation.

The BIOCLIM program contains a number of climatic surfaces covering the Australian continent. For each point within a given area, BIOCLIM can generate an interpolated value for up to 35 different parameters. These are:

1. Annual Mean Temperature
2. Mean Diurnal Range(Mean(period max-min))
3. Isothermality 2/7
4. Temperature Seasonality (C of V)
5. Max Temperature of Warmest Period
6. Min Temperature of Coldest Period
7. Temperature Annual Range (5-6)

8. Mean Temperature of Wettest Quarter
9. Mean Temperature of Driest Quarter
10. Mean Temperature of Warmest Quarter
11. Mean Temperature of Coldest Quarter
12. Annual Precipitation
13. Precipitation of Wettest Period
14. Precipitation of Driest Period
15. Precipitation Seasonality(C of V)
16. Precipitation of Wettest Quarter
17. Precipitation of Driest Quarter
18. Precipitation of Warmest Quarter
19. Precipitation of Coldest Quarter
20. Annual Mean Radiation
21. Highest Period Radiation
22. Lowest Period Radiation
23. Radiation Seasonality (C of V)
24. Radiation of Wettest Quarter
25. Radiation of Driest Quarter
26. Radiation of Warmest Quarter
27. Radiation of Coldest Quarter
28. Annual Mean Moisture Index
29. Highest Period Moisture Index
30. Lowest Period Moisture Index
31. Moisture Index Seasonality (C of V)
32. Mean Moisture Index of High Qtr. MI
33. Mean Moisture Index of Low Qtr. MI
34. Mean Moisture Index of Warm Qtr. MI
35. Mean Moisture Index of Cold Qtr. MI

Using BIOCLIM, for each point in WABiota the above parameters were evaluated and summary statistics calculated. A program was written at the WA Herbarium that took these BIOCLIM outputs and tested the extent to which each individual point's climatic parameters were within the environmental envelope described by the points as a whole.

While there are a number of robust approaches for testing for outliers (Digby & Kempton 1987, Gauch 1982, Palmer 1993), the method adopted here was to test whether a given climatic attribute was more than two standard deviations outside the mean of that parameter. This simple approach is consistent with the levels of positional accuracy likely to be found for the many historical records which are the most problematic area requiring geocode validation.

The total number of climatic attributes within acceptable limits was summed. This total was then classed into four main categories (Table 2):

Table 2. WABiota Validation Categories

Class	Number of parameters outside two standard deviations	Interpretation
1	0	Validation passed. Point is considered as acceptable and no further validation is required.
2	1 – 10	Validation passed. While some parameters were outside acceptable limits, sufficient were inside the limits to not warrant further validation.
3	11- 20	Validation failed. While point may be legitimate, it is of sufficient cause for concern to warrant further investigation.
4	20 – 35	Validation failed. Point is highly likely to be in error and should be investigated.

The above classification simplifies a complex situation. It does not consider whether some parameters strongly influence distribution for a given organism more than others. Ideally these parameters should be weighted to reflect those influences. However, such an exercise would necessarily be specific to particular groups. Given the large number of taxa being tested, the above classification would still yield sufficiently valuable information across the entire range of taxa to target those records most in need of checking.

A number of other criteria were used in determining validation class, such as whether a point was mislocated over an ocean or outside the predictive area of the BIOCLIM surfaces.

A database was generated containing a validation class for each record within WABiota.

Those records with classes one and two were not validated any further. Those records in WAHerb that failed validation were then checked manually and WABiota was updated to reflect the validated co-ordinates. Records from other data sources were not validated any further as this was beyond the resources of the project.

Geocode validation of vouchered specimen data in WAHerb entailed comparing the existing geocode entry with one derived from a re-evaluation of the specimen label locality statement. An ARC/Info interface to the WAHerb validation database was created to aid in this comparison.

Because of slight discrepancies between the coverages for coastline and the digital elevation model (DEM), many specimen records close to the coast were erroneously reported as non-valid. Priority was given to the validation of records clearly within or beyond the coastline.

A total of 5,086 records were validated in this manner, 1,752 existing geocodes were validated as being correct, and 3,334 records required a new geocode to be calculated. These data were then used to update the main WAHerb database records.

A proportion of the records where the existing geocode had been found to be correct was then checked for inaccuracy of taxonomic determination as the prime cause for failing the initial validation. After re-determination and updating of the WAHerb database records, a final refresh of the WABiota data warehouse was performed.

2.4 Species Prediction Modelling

Preliminary investigations indicated that large portions of the RFA area were under-represented within existing databases. It was also the case that particular areas, being the subject of research, contained large numbers of records in comparison to other areas, resulting in extreme sampling biases. An advantage of using predicted distributions is that not only can they account for potential distribution in areas where gaps may exist, but also compensate, to some extent, for the effects of high unit collector effort.

SpModel, a software package developed for Environment Australia (Ferrier & Watson 1996), was used to generate predicted distributions for each species within WABiota. Only indigenous, vascular flora with current nomenclature were modelled. SpModel required a minimum of ten points for any given species. This reduced the candidates for modelling from over 3,000 species to 1,693 species. In the event of concern that the ten point minimum might exclude a large number of threatened taxa with limited populations, it should be noted that these taxa are generally far better collected than non-threatened taxa.

The following discussion examines the various choices confronting the selection of an appropriate modelling environment. The primary factor in determining these choices was a substantial number of species to model and limited time in which to process them.

SpModel, based on the statistical package Splus™ (Mathsoft, 1995), had the capability to model using GLM (Generalised Linear Model) and GAM (Generalised Additive Model). Whilst GAM had been stated as providing a superior model to that provided by GLM (Ferrier & Watson 1996) it required greater computing resources, resulting in longer execution times.

For this exercise, the GLM model was chosen over GAM for a number of reasons. Firstly, the data to be modelled were treated as primarily presence-only data due to the large proportion of opportunistic records within a number of the data sources. Secondly, many of the historical records required caution in relation to geocodes. The superior modelling of GAM might have been compromised by variability in spatial accuracy of a substantial proportion of existing available data. (Despite these concerns, initial testing of a range of species showed the GAM model to have slightly better predicted distributions.) Finally, the increased requirement of processing time by GAM over GLM would have substantially extended the processing of so many taxa. For these reasons, GLM was adopted as an acceptable compromise between processing time and confidence in the model.

SpModel also had the capability to model geographic space, a method for identifying clusters of points. However, this factor greatly increased computational time and for this reason it was not included as a predictor. Interestingly, using geographic space as a predictor failed on test models using a number of species. The cause of this failure was unknown.

Potential spatial predictor variables for SpModel included continuous surfaces such as climatic attributes or categorical variables such as geology or vegetation. Site variables such as collection date could also be included to test for relationships between those variables and existing distribution though they would not influence the predicted distribution.

A range of predictor variables were assessed for their potential use within the modelling process. While there were 35 climatic parameters and a few categorical variables such as geology and vegetation available for inclusion, the incorporation of all of them would have unacceptably increased processing time. A suitable subset of climatic parameters was chosen that incorporated temperature, precipitation and radiation levels at both annual and seasonal levels. The variables chosen for use by the model were:

Annual Mean Temperature
Annual Precipitation
Maximum Temperature of Warmest Period
Minimum Temperature of Coldest Period
Annual Mean Radiation
Annual Moisture Index
Precipitation of Wettest Period
Precipitation of Driest Period
Mean Temperature of Wettest Quarter
Mean Temperature of Driest Quarter
Mean Temperature. of Warmest Quarter
Mean Temperature of Coldest Quarter
Precipitation of Wettest Quarter
Precipitation of Driest Quarter
Precipitation of Warmest Quarter
Precipitation of Coldest Quarter
Aspect
Slope
Geology (Regolith + Precambrian)
Forest Ecosystems
Ecological Vegetation Systems

The Forest Ecosystems grid documented a classification of forest ecosystems within south-western Australia that was developed as part of the CRA (Anon, 1998). The Ecological Vegetation System grid was a higher level classification of the Vegetation Complex database compiled by Mattiske Consulting. A soil database commissioned for the RFA became available close to the conclusion of this project. It was expected that soil data would provide a strong correlation with species distribution (E.M. Mattiske pers. comm., Havel 1968, 1975b). However, initial investigation of the soils data showed legend information to be substantially incomplete, making verification of the data difficult. Consequently, the soil map was not used in the modelling process.

The above grids were generated on a longitude/latitude base at a resolution of 9 seconds of arc. This resolution was determined by the resolution of the DEM for the RFA area. A DEM was required by BIOCLIM to generate the climate grids.[2]

For this exercise no site variables were included in the modelling as they unnecessarily slowed down the run-time for each model for no benefit to the predicted distribution.

The output of SpModel for presence-only data was a likelihood grid. Each cell within the grid has an index ranging between zero and one, with zero indicating no likelihood for the presence of the modelled taxon and with one indicating extremely high likelihood of the taxon being present at that location. The likelihood index should not be confused with a

[2] The source climatic surfaces in BIOCLIM are recorded at a resolution of one fortieth of a degree. These values are then interpolated onto the resolution of the DEM.

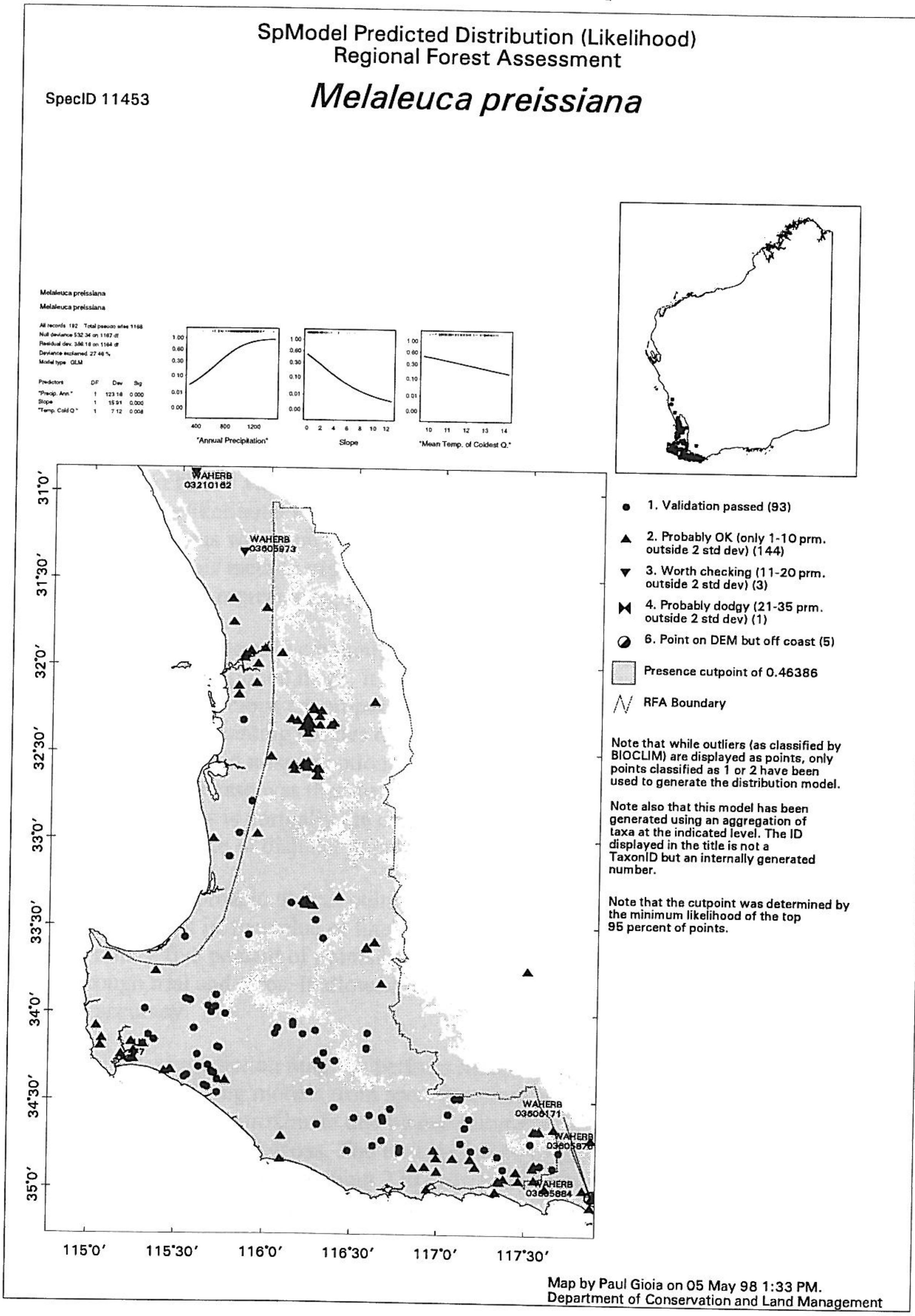
probability index which carries a much greater strength of prediction but is only generated when using presence/absence data (Bennet, Watson & Barratt, 1997).

A likelihood grid was generated for each current, indigenous taxon within the RFA area having at least ten records. Initial results indicated that taxon nomenclature should be aggregated to species level due to the number of records in WABiota identified as species and not as sub-species or varieties. In any case, the intent of the project was to contribute to the understanding of diversity at the species level.

It was acknowledged that some sub-taxa with disjunct distributions might be merged together. However, given the known under-representation of many taxa within the RFA area, it was hoped that models would benefit from the addition of extra points and that the distributions of most sub-taxa would not excessively influence the overall result. This must also be viewed in light of the large number of taxa being merged together, that the contribution of individual models with over-predicted distributions would become less significant.

Only records validated to the second level of reliability by BIOCLIM (see Table 2) were used to generate these models. A map was produced for each species displaying the extant distribution in WA, the predicted distribution within the study area shaded by increasing value of likelihood, the ANOVA table for the GLM statistic and a legend with symbols describing the levels of BIOCLIM validation for the species data (Figure 1).

Figure 1. Example of predicted distribution of *Melaleuca preissiana*



Each map was examined using the above information to assess whether the predicted distribution model was realistic. Unacceptable prediction maps, usually those over- or under-predicting possible species distribution by more than 30% were annotated in a database. These maps were excluded from species richness counting.

Experienced botanists were asked to validate the prediction maps for families or genera with which they were familiar. These botanists included Dr. Neville Marchant (*Agonis, Chamelauceum, Darwinia, Drosera*), Bruce Maslin (*Acacia*), Andrew Brown (Orchidaceae), Dr. Terry Macfarlane (*Amphipogon, Austrostipa, Stipa, Lomandra*), Ray Cranfield (various taxa) and Mike Hislop (various taxa).

Models were also compared against independently compiled distribution maps for about 150 taxa published in Churchill (1961).

The species distribution likelihood maps generated by SpModel were, in themselves, insufficient for deriving an index of species richness. The likelihood values generated for each species were highly dependent on the strength of correlation between source data and predictor variables. Before a species count could be performed, a method for standardising likelihood values across all species was required. We tested a method that obtained the mean likelihood for all points for a given taxon, using that mean as a cutpoint below which locations would be tagged as species absent and above tagged as species present. However, this method proved unsatisfactory, with most maps either over-representing or under-representing true distributions by an unacceptable amount.

When observing how the above botanists assessed the suitability of a predicted distribution from the supplied maps, it was noted they generally selected a cutpoint resulting in a range envelope encompassing the majority of points. This process was akin to the practice of representing species distributions as "blob" maps, rather than points, commonly done to avoid misperceptions regarding true distribution due to collector bias. The difference in this case was that the range envelope was calculated statistically, rather than drawn by hand as is normally the case. (A point of caution was noted here in regard to the possible presence of disjunct populations.)

Using this simple approach, the cut point was set at likelihood value corresponding to the fifth percentile. That is, the value of likelihood at which five percent of points had a lower value and ninety-five percent of points had a greater value. This percentile was arrived at largely through trial and error. It allowed for a small proportion of points having lower positional accuracy.

A species richness prediction map, or, perhaps more correctly, a cumulative likelihood map, was generated using models from species annotated as acceptable. The unit of counting was set at an approximate one by one kilometre square (100ha). This value was arrived at for a number of reasons. The base data used to generate the predicted distribution was nine seconds, or about 270 metres. The lower limit for the unit of count should be greater than 270 metres by at least two or three times to accommodate fluctuations in accuracy. The upper limit was determined by the minimum level of accuracy required by the integration and planning phases of the RFA process.

The smallest unit of operation should correspond to the size of reserves being considered. The variation in areas of existing CALM estate was examined, taking into consideration the large number of small reserves, with area less than 100 ha. Thus, the unit of counting should be the smallest value possible to approach this level of accuracy. This resulted in the adoption of a one by one kilometre unit of counting.

3 Results

3.1 WABiota

A total of 153,212 vouchered and non-vouchered flora records were warehoused in WABiota from a range of sources internal and external to CALM (see Tables 3 and 4 below). These records comprised 3,244 current, indigenous taxa with at least one record within the RFA area.

Table 3. CALM Sources

Source	Vouchered	Records
WAHerb Original Records	Yes	36137
Regional Forest Agreement 1997 –Ecological Vegetative Systems	Yes	3954
Regional Forest Agreement 1997 – Key Ecological Attributes	Yes	9392
Floristic survey of the Tingle Mosaic (Wardell-Johnson *et al.* 1989, Wardell-Johnson *et al.* 1995)	Some	15056
Havel site-vegetation type bulletins (Havel 1975a, Havel 1975b)	No	5126
CALM Declared Endangered Flora Database	Some	2949
Banksia Atlas (Taylor & Hopper 1988)	No	2896
Total		**75510**

Table 4. External Sources

Source	Records
Alcoa of Australia Limited	19784
Worsley Alumina Pty Ltd	9864
Scott River National Park survey for BHP (Mattiske Consulting 1996)	637
Griffin Coal Mining	1985
Water and Rivers Commission and Water Corporation	9862
John Forrest National Park and Red Hill Survey (Mattiske & Burbidge 1991)	745
Mt Westdale – Dobaberry Swamp (Trudgen 1984)	813
Per Christensen PhD data (Christensen 1980)	4662
RFA studies by Mattiske Consulting (Mattiske Consulting 1997a)	27334
Shire of Mundaring (Mattiske Consulting 1997b)	2016
Total	**77702**

3.2 Species of Special Interest

Species attribute tables were used to select points within various JANIS target groups. These groups included DRF, poorly known taxa, relictual taxa, species endemic to the RFA area within a 150 kilometre boundary and taxa at their range ends or having disjunct populations.

The distribution of DRF and priority flora was included in the final CRA report (Gioia et al., 1998).

3.3 Species Prediction Modelling

Distribution models were generated for 1,693 current, indigenous species with at least one record within the RFA area.[3] Distribution maps were printed for each species and assessed for suitability. After assessment, 263 models were rejected because the model was not considered to sufficiently reflect true or likely distribution, leaving a total 1,430 acceptable models.[4]

The most common predictors for distribution were climatic surfaces, followed by slope and geology. The most common climatic surfaces employed by the model were those incorporating some aspect of seasonality in temperature and precipitation. Only very occasionally were forest ecosystems used by the model.

The number of species within one by one kilometre grids were counted across the entire RFA area. Counts for all taxa and for endemic taxa were generated and maps at 1:500000 were printed. Maps including all taxa and those for just endemic taxa were generated.

A number of areas suggesting high species richness were revealed: the Whicher Range and the Scott River Plains in the south-west of the RFA study area, the Blackwood River plateau, an area north-east of Walpole in the south-east of the RFA study area and, to a lesser extent, the Kalamunda-Mundaring area east of Perth in the north of the RFA study area.

The species richness prediction maps were checked in a number of ways. Raw data from WABiota were counted for sections of the map and found to support the predictions of high species richness. Data counted for some areas of low species richness were much lower than the predicted values, particularly in the Collie basin, probably due to under-sampling. Experienced botanists Neville Marchant, Greg Keighery, Libby Mattiske and Roger Hearn confirmed that a number of areas predicted by the map to be highly species rich were consistent with their knowledge of those areas, though all found very high species richness around the Blackwood River plateau to be somewhat unexpected. They also confirmed that the results for the Collie basin under-represented likely richness in that area.

The predicted centres of endemism using predicted distributions largely concurred with centres of endemism generated by Roger Hearn using an alternative method (R. Hearn, pers. comm.). This method made use of the Vegetation Complex map (Mattiske Consulting Pty Ltd, 1997a) by overlaying vegetation complexes with source point data and counting the number of taxa restricted to a given vegetation complex. This process included all records of DRF and poorly known taxa, in contrast to the predictive modelling process which only modelled species with ten or more populations. This alternative method demonstrated similar richness trends to the SpModel results, including the Blackwood Plateau, with the qualification that richness within the Wicher Range and associated ironstone communities, and the Scott River Plain, was higher than that calculated using the SpModel process. This might be partially explained by the relatively large number of DRF and poorly known species thought to occur in those areas, which were included in the alternative method but reduced in number by the SpModel process.

---

[3] Records were aggregated at the species level for processing by SpModel.

[4] In the publication Anon (1998), this figure was incorrectly cited on page 131 as 1929. That figure included non-current and naturalised species. The correct number of accepted models for current, indigenous species was 1430.

A species richness map for the RFA area was included in the final CRA report (Gioia et al., 1998).

4 Discussion

The production of "predicted" species distributions, and the subsequent calculation of a species richness index using those distributions, raise a number of issues requiring clarification if the richness map is to be interpreted correctly. Only when these issues are clearly understood and addressed by potential users, will the technologies that make these outputs possible find a useful place in biodiversity assessment and the determination of priorities for future field survey.

A number of questions come to mind regarding the species modelling process. What do the predicted distributions actually represent? What ecological reality, if any, does the species richness map portray? What error factor is associated with these outputs? What scale should they be used at?

In the first instance, the predicted distributions (and their limitations thereof) should be well understood. It has been stated previously that the most common predictor variables were climatic. Only occasionally were predictors such as forest ecosystems, geomorphology or slope selected. This was somewhat unexpected. It was also noted that continuous variables, primarily, provided the best correlations with source data points.

One possible explanation for this, among others, is related to the spatial accuracy of species geocodes. It was known that a number of historical records had variable spatial accuracy, ranging from a few kilometres to, in some instances, tens or hundreds of kilometres. When overlaying a species geocode with a continuous variable, such as rainfall, spatial error of a few kilometres is unlikely to dramatically affect the corresponding climatic values which themselves change only gradually, at a regional scale. This is particularly true in the context of the RFA area having relatively little relief to affect climatic parameters, certainly in comparison to the eastern states of Australia.

However, in the case of categorical variables, this is not necessarily the case. If the spatial error of a point approaches or exceeds a given proportion of the size of polygons representing the categorical variable, the actual polygon the point falls in may be random. In this situation, the modelling software would be unlikely to detect a strong correlation, or one that was not due to chance.

It may also be that the categorical variables employed were not sufficiently meaningful in the context of the species analysed to form any adequate correlation.

In any case, the vast majority of predicted species distributions were based solely on climatic attributes, which themselves were interpolated from an initial resolution of 1/40th of a degree. Therefore we cannot expect these results to reflect distributions where limiting factors are determined by micro-environmental situations. Some examples might include plants favouring water-gaining sites, particular soil types, aspect, or micro-climatic effects such as west-facing river headways (N.G. Marchant, pers. comm.). *Any products derived from these maps should therefore be assessed at a regional scale only.*

A specific case in point is that cited above in Figure 1, *Melaleuca preissiana*. The predicted distribution reflects the known extent of the species (Barlow & Cowley, 1988). However, internally, it overstates the true distribution of *M. preissiana*, which only occurs in wetter areas.

Another point to bear in mind is that the predicted distributions do not take into account human disturbance such as clearing. In some cases, species are predicted to occur in areas with no records. One example of this is an area south of Bunbury on the Swan Coastal Plain. This area was cleared very early in the settlement of Western Australia and consequently provides very few historical records of the biota prior to clearing. Yet a number of species were predicted to occur in this area. Similarly for the Whicher Range, which has also been substantially cleared.

Because climatic factors were the most common range determinants, for the reasons cited above, most of the predicted distributions will be overstatements of true distribution. However, this is offset, to some extent, by the use of presence-only source data which often understate true distribution through inadequate or biased sampling.

Another issue requiring clarification is the extent to which the overlaying and summing of predicted distributions produces a meaningful index of species richness. In a number of well-collected areas already known to be species rich, there was an apparent concurrence with areas predicted to be species rich using predicted distributions. These areas included the Whicher Range, Scott River National Park and an area to the northeast of Walpole. This concurrence seemed to be supported by an analysis of the raw data alone. However, there were other areas where there seemed to be a large discrepancy between the predicted figures and the raw data, particularly in the Collie basin and an area to the east of the Collie basin, north of the Muir Highway. It is highly likely that this discrepancy is the result of poor collecting in that area. Thus, while the predicted distributions have been demonstrated to be an improvement on using raw data alone, they still do not compensate completely for poor collecting.

Where sampling has been adequate, areas suggested as being species rich, which are not already recognised as such, warrant on-ground validation to confirm their existence. Other areas where there is a discrepancy between predicted and recorded data, such as the Collie basin and north of the Muir Highway, which may be the result of inadequate sampling, also warrant further survey to provide a more informed basis for decision making. In this context, a data review using software such as DAM, in conjunction with the species richness index, would greatly aid the process of identifying and prioritising areas for future survey.

In the species richness map published in Gioia et al. (1998), figures for species counts were displayed as contours overlaying a shaded richness index. When citing these figures, it is important to remember they are generated from predicted distributions transformed from a continuous likelihood grid to a binary presence/absence grid. Whilst the absolute value of species counts may concur with true richness in some areas, they must still be regarded as statistically determined. Consequently, there may be an argument for not citing these actual values, instead normalising them to a single index indicating high or low richness, or perhaps not citing them at all. One problem with this approach is the loss of a sense of magnitude. As an extreme example, an area with few species overall might still be depicted with areas of high richness even though very few species might be involved. Contrastingly, a region highly species rich across a large proportion of its area might have only those areas with the highest richness highlighted, even though the entire area was relatively species rich. On balance, it is preferable to cite actual values, on the understanding that these are predicted values of likelihood.

One point of ambiguity which needs to be addressed is the question of resolution. Although the predicted distributions were resampled from a polygon to grid base of one by one kilometre, approximately, this does not imply the predicted distributions were accurate to that level, remembering that the most common predictors were climatic grids which were originally sampled at 1/40th of a degree. The primary reason for adopting this approach was to retain the shape of the distribution rather than generalising it to a much larger block, as is more commonly the case. One advantage of generalising to a large block, say ten by ten minutes of arc, is that the resolution of the prediction becomes immediately apparent, and is interpreted accordingly. However, this occurs at expense of retaining some environmental meaningfulness to the shape of a given region of high species richness. In the case of the Blackwood plateau, it becomes immediately obvious that the region of high richness around that plateau is somehow related to features of the plateau, be they climatic or otherwise. This might be totally obscured by adopting large blocks of resolution, depending on their size and the point at which their edges occur in the landscape. On balance it seems preferable to retain the morphological relationship between a zone of predicted richness and the underlying area. However, such a map should be accompanied by an explanation its regional applicability.