JOINT VEGETATION MAPPING PROJECT

Brigalow Belt South

Stage 2

NSW WESTERN REGIONAL ASSESSMENTS

FINAL REPORT February 2004

Resource and Conservation
Assessment Council

JOINT VEGETATION MAPPING PROJECT

BRIGALOW BELT SOUTH

WESTERN REGIONAL ASSESSMENT

STAGE 2

A cooperative project coordinated by
The Resource and Conservation Division
Office of Sustainable Development, Assessments and Approvals
Department of Planning, Infrastructure and Natural Resources (DIPNR)
(formerly PlanningNSW)

in conjunction with partner agencies:
NSW Department of Land and Water Conservation (now DIPNR)
NSW National Parks and Wildlife Service (now Department of Environment and Conservation)
State Forests of NSW

Undertaken for
The Resource and Conservation Assessment Council (RACAC)
NSW Western Regional Assessments

Project number WRA 24

For more information and for information on access to data contact the:

Resource and Conservation Division,
Office of Sustainable Development, Assessments and Approvals
DIPNR
PO Box 39
SYDNEY NSW 2000

Phone: (02) 9228 6586
Fax: (02) 9228 6411

ISBN 1 74029 184 0

This project was coordinated by the Resource and Conservation Division (RACD), Department of Planning, Infrastructure and Natural Resources (DIPNR), formally PlanningNSW.

In-kind funding was provided by NSW Department of Land and Water Conservation (now DIPNR), NSW National Parks and Wildlife Service (now DEC), PlanningNSW (now DIPNR) and State Forests of NSW (SFNSW).

Project Coordinator: Peter Lezaich, RACD.

Project management: The Joint Vegetation Mapping Project (JVMP) Management Committee comprising representatives from each agency: Angela McCauley and Gary Saunders (DEC), Pat Groenhout and Doug Binns (SFNSW), Bob Denholm and Greg Steenbeeke (DIPNR) and Peter Lezaich (Project Coordinator, DIPNR).

Technical supervision and writing: The JVMP Technical Working Group comprising representatives from each agency: Peter Lezaich (Project Coordinator), Angela McCauley (DEC), Doug Binns (SFNSW), Greg Steenbeeke, Dominic Sivertsen and Ross Peacock (DIPNR).

Publication assistance: Ric Noble (editing coordinator), Tracie Bird-Gardiner (editing and map production), Bronwyn Rice (report format), Ray Farley (independent editor), and an anonymous referee from CSIRO.

Additional thanks to Simon Ferrier and Glen Manion (DEC GIS Research and Development Unit), Vanessa Pelly, Ben Scott, and Susan Rhind (DEC Western Regional Assessments Unit).

CONTENTS

Table of Figures

List of Tables

1. PROJECT SUMMARY

This report describes a project undertaken for the Resource and Conservation Assessment Council as part of the regional assessments of western New South Wales. The Resource and Conservation Assessment Council advises the State Government on broad-based land use planning and allocation issues.

1.1 PROJECT OBJECTIVES

The key objectives of the project were to provide:

1. an extant vegetation map, which showed the distribution of modelled vegetation groups at the landscape level;

2. a pre clearing vegetation map, which showed the potential distribution of modelled vegetation groups at the landscape level;

3. six map sheets completed to Department of Land and Water Conservation (DLWC) Native Vegetation Mapping Program (NVMP) technical standards as set out in the Guidelines for mapping native vegetation (Sivertsen and Smith, 2001);

4. floristic classification of all sampled native plant species identified via systematic floristic survey within the BBS bioregion using agreed classification techniques.

1.2 METHODS

The Joint Vegetation Mapping Project (JVMP) utilised the following methods for mapping the vegetation groups across the landscape:

1. Data audit and gap analyses were used to determine the priority locations for full floristic survey.

2. Full floristic survey of the BBS bioregion incorporated floristic, physiographic and structural information that met the DLWC *Guidelines for mapping native vegetation*.

3. Aerial photographic interpretation (API) was carried out to the technical standards as described in the *Guidelines for mapping native vegetation*. API was carried out as either full NVMP spatial extent or as targeted API to the agreed technical standard.

All available API data were then compiled to produce a composite API vegetation layer for use in the modelling process.

4. Data were entered into the NPWS Vegetation Survey Database (NPWS, 2002) and NPWS YOWIE database. These relational databases provide ready access to data and utilise the Microsoft Access database platform. This allows ready use of the queries, forms, reports, macros and modules allowing easier interrogation and interpretation of the data.

5. Data preprocessing was required before analysis so that the various floristic survey data sets could be brought to a uniform standard. Primarily standardisation of the Braun-Blanquet scale for cover abundance was required along with substantial changes to the taxonomic tables to ensure nomenclature was current.

6. Data analysis was carried out using PATN software to investigate the relationships between survey sites and floristics. PATN encompasses a suite of multivariate statistical tools which utilise both hierarchical and non-hierarchical methods. One hundred and fifteen vegetation groups were identified, across the BBS bioregion, utilising this process.

7. Data modelling was carried out by the NPWS GIS Research and Development unit in Armidale with technical input and post modelling analysis by the JVMP Technical Working Group (TWG). A Generalised Dissimilarity Model or GDM was utilised to develop the relationship between a suite of edaphic variables (which relate to soil) and the floristic survey data.

The vegetation groups as derived from the PATN analysis were then introduced to the model to determine the relationship between the modelled space and the vegetation groups. The composite API vegetation layer was introduced to the model to act as a constraint on the model.

1.3 KEY RESULTS AND PRODUCTS

Key products:

1. The production of 115 probability surfaces providing information on the potential distribution of vegetation groups within the bioregion, suitable for landscape level planning at a bioregional scale.

2. The production of 115 probability surfaces masked to the extant vegetation; showing the current distribution of native vegetation groups within the bioregion, suitable for landscape level planning at a bioregional scale.

3. The production of a composite map showing the combined potential distribution of vegetation derived from the modelled probability surfaces suitable for landscape level planning at a bioregional scale.

4. The production of a composite extant vegetation map, based on the potential vegetation composite map and the DLWC land use data set, showing the current distribution of native vegetation across the bioregion, suitable for landscape level planning at a bioregional scale.

Key results:

1. Increased level of knowledge of the vegetation community-environment relationships and of the floristic diversity of the vegetation groups of the BBS bioregion.

2. Modelling of 2 739 814 hectares of extant native vegetation, which accounts for 52% of the area of the bioregion, presented in mapped form.

2. INTRODUCTION

2.1 BACKGROUND

2.1.1 Western Regional Assessment

The Western Regional Assessment (WRA) process was implemented within the Brigalow Belt South (BBS) bioregion in 1999. The WRA process was initiated by the NSW Government to gather information at a regional scale to assist in the formulation of management strategies for the public lands within the BBS bioregion.

The BBS bioregion assessment process was implemented in two stages. Stage one was completed in February 2000 and was concerned with the assessment of state forests, national parks and vacant Crown land south of Narrabri within the BBS bioregion (Figure 1). Stage 2 projects were implemented in late 2000 and the JVMP was approved in July 2001. The JVMP was designed to fill in the gaps in vegetation survey and mapping by expanding the Stage 1 assessment to include all identified native woody vegetation throughout the BBS bioregion across all land tenures. The JVMP was conducted according to the technical standards adopted by the Department of Land and Water Conservation (DLWC), for their Native Vegetation Mapping Program, and detailed in the DLWC Guidelines for mapping native vegetation (Sivertsen and Smith, 2001).

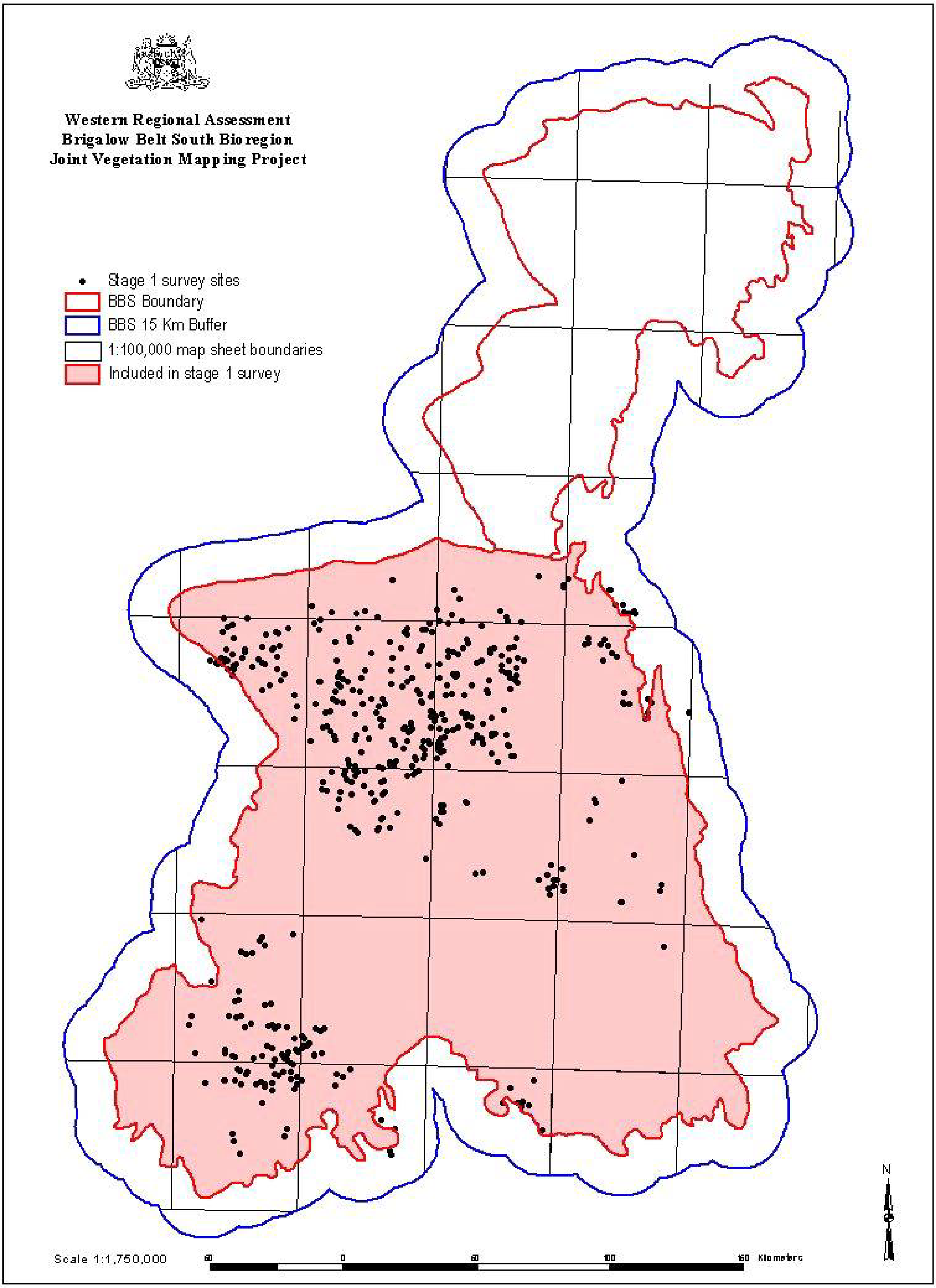
The project was coordinated by the Resource and Conservation Division (RACD). It utilised expertise and staff from the following agencies: RACD and PlanningNSW (DIPNR), DLWC (DIPNR), NPWS (DEC), and SFNSW (DPI). The project was conducted under a partnership agreement between the agencies.

2.2 TECHNICAL WORKING GROUP AND MANAGEMENT COMMITTEE

The JVMP was conducted under the direction of a Management Committee which initially met on a monthly basis then as required while maintaining input and oversight via e-mail communication.

The technical aspects of the JVMP were directed by the Technical Working Group, which comprised staff members with vegetation mapping expertise from each of the partner agencies. Meetings were held at a minimum of once each month. The TWG was responsible for all aspects of technical decision making throughout the project.

Figure 1: Extent of BBS bioregion Stage 1 vegetation surveys



2.3 PROJECT OBJECTIVES

2.3.1 Extant vegetation model and map

To produce a model and a map of the extent of the current distribution of extant native woody vegetation, using agreed techniques, based upon comprehensive floristic sampling and strategic aerial photography interpretation across the Brigalow Belt South bioregion.

2.3.2 Preclearing vegetation model and map

To produce a preclearing vegetation model and map of the BBS bioregion, using agreed techniques, based upon comprehensive floristic sampling and aerial photography interpretation of current extant vegetation across the BBS.

2.3.3 Native Vegetation Mapping Program map sheets

To map and floristically classify six agreed 1:100 000 topographic map sheets within the BBS bioregion in accordance with the DLWC NVMP *Guidelines for mapping native vegetation*.

2.3.4 Floristic classification of all sampled plant species

To floristically classify all sampled plant species identified via systematic floristic survey within the BBS bioregion using agreed classification techniques.

2.4 PROJECT OUTPUTS

2.4.1 Extant vegetation model and map

The JVMP produced an extant vegetation model, presented in mapped form, for the BBS bioregion. The map shows the modelled distribution of extant vegetation, identified through aerial photography and satellite image interpretation. The extant vegetation model was restricted to the extant woody vegetation with an approximate crown canopy projection of greater than 10% and identified open woodland / grassland vegetation groups. The extant vegetation model was designed as a tool suitable for landscape level planning at a bioregional scale.

2.4.2 Preclearing vegetation model and map

In the BBS bioregion the concept of a preclearing map was difficult to precisely define and its meaning was therefore unclear. Anthropogenic influences, particularly in the last 200 years, have influenced landscape processes and may have irreversibly altered some elements of landscape function, though to what extent is unknown. Levels of disturbance vary throughout the bioregion and result from a myriad of management regimes, as is typical in an agrarian landscape.

Preclearing vegetation models are commonly based on the current extant distribution of vegetation, as determined through a sampling regime (Jorgensen, 1994). Through such sampling, vegetation composition, structure and distribution may be determined and then modelled to represent a preclearing landscape (Smith 2000). However, in such cases the model's relevance to preclearing vegetation is rarely tested (Oliver et al, 2002; Smith, 2000) and its validity is therefore unknown.

RACAC (1999) indicated that geographically restricted and/or highly degraded sites were likely to be underrepresented in model datasets and therefore have lower levels of accuracy than for sites which had a widespread distribution or were relatively common. Further changes to the vegetation within the landscape may occur through the application of different management regimes. Lunt (1997) demonstrated how the changes to the grassy forests and woodlands of the Gippsland Plain have resulted in two different vegetation communities, where once there existed a single community.

It was not possible to produce a preclearing map which meets the definition of the Native Vegetation

Conservation Act and which could be supported through documented data based on independent validation of the correlation between extant and preclearing vegetation distribution. For these reasons the JVMP was unable to produce a preclearing model and map.

The JVMP has instead produced a "predicted potential vegetation distribution model" that showed the potential for a particular vegetation group, expressed as a probability, to occur in a given environment. The potential distribution is dependant on the current extant distribution, as defined by the sampling regime, and the relationships such vegetation groups have with the suite of environmental factors utilised by the model. The predicted potential vegetation distribution model was designed as a tool suitable for landscape level planning at a bioregional scale. This product has substantial utility in land repair and revegetation projects and is designed as a tool for looking forward rather than backwards.

2.4.3 Native Vegetation Mapping program map sheets

The JVMP has produced data for five map sheets to NVMP standards as described in the DLWC *Guidelines for mapping native vegetation v2.1* (Sivertsen and Smith, 2001). Aerial photography interpretation (API) and floristic sampling were carried out to NVMP technical standards for the following 1:100 000 map sheets: Gravesend, Curlewis, Boggabri, Tambar Springs and Coonabarabran.

All NVMP data, from floristic surveys, was entered into the NPWS Vegetation Survey Database (NPWS, 2002) and YOWIE database and included in the complete BBS bioregion dataset. This floristic sampling and API data was utilised by the JVMP and the resulting vegetation groups were derived from the complete BBS bioregion data set. The NVMP map sheets were subject to a separate project report and will not be dealt with individually in this report.

2.4.4 Floristic classification of all sampled plant species

The JVMP has produced a floristic classification of all native plant species identified during the survey process. One hundred and fifteen vegetation groups were identified, through multivariate analysis, as being likely to be found within the BBS bioregion. A further three vegetation groups were derived through the API program. Probability surfaces were derived for each of the likely vegetation groups. Areas within the landscape were identified as having a level of probability of occurrence ascribed to each vegetation community, as assigned by their environmental and geographic features.

2.4.5 Note on mapping scale

The models produced are applicable to bioregional planning projects only and are not suitable for property-scale planning or mapping.

3. METHODS

Vegetation mapping processes have typically included a number of different components combined to produce the final products. The JVMP had eight primary components:

- gap analysis
- floristic survey
- aerial photography interpretation
- data entry
- data preprocessing
- data analysis
- data modelling
- product integration.

Some of these components produced primary data whilst others were derived from the primary data sets.

The BBS bioregion was extended by the addition of a 15 kilometre buffer zone. This buffer zone was introduced to allow the use of pre-existing data, which occurred within the buffer, to be utilised in the modelling of the vegetation distribution. This allowed vegetation groups with limited distribution within the BBS bioregion and which were more widely distributed in five neighbouring bioregions to be identified. The buffer was included in all data analyses, modelling and product integration for the JVMP. Mapped outputs are presented only to the BBS bioregion boundary, as land within the buffer may be subject to other assessment processes (for example, State Biodiversity Assessment, Nandewar Bioregional Assessment). For the purpose of this report, the BBS bioregion refers to the BBS bioregion and the 15 kilometre buffer combined.

3.1 GAP ANALYSIS

Gap analysis was used to determine the adequacy of existing survey data in relation to the environmental and geographical space in which survey sites occurred. Through gap analysis, survey design was improved by focussing attention on those areas within the landscape which produced the greatest quality of information for a specified amount and within a specified timeframe.

3.1.1 The survey gap analysis tool

The survey gap analysis tool was developed by the NPWS GIS Research and Development Unit in Armidale. The survey gap analysis tool is an Arcview GIS extension written in the computer programming languages of Avenue script and C++. The tool was first developed for the North East Comprehensive Regional Assessments and is summarised in Appendix 11.

The gap analysis tool was used to select survey sites that representatively covered environmental and geographic gradients occurring throughout the study area. The underlying principle is that the survey coverage is analysed directly in relation to the underlying continuous environmental and geographic space rather than being an arbitrary categorisation of such space (as is often the case following traditional stratified sampling methods). Existing floristic plots are considered when selecting the location of new sites so that the survey effort is maximised.

The tool randomly generates a set of candidate sites (for example: 10 000 and 20 000 as used by the JVMP) from which the target survey sites are selected. The target sites aim to sample the environmental space (as represented by the candidate sites) not sampled by previous surveys. In this case, abiotic layers were weighted according to expert opinion; the level of importance of each layer was ranked and used in determining priority survey areas (rather than using equal weighting).

Site selection can be an ongoing, reiterative process. As the vegetation surveys are completed, those site locations can be added to an existing “sites database” and the gap analysis tool rerun.

3.1.2 Existing plot data

A data audit of existing floristic survey data was carried out to determine the availability and effectiveness of such data to the JVMP. A total of 12 survey databases were accessed containing 33 individual floristic surveys that included work in the BBS bioregion and met the requirements of the JVMP sampling strategy and criteria. In total 1 922 existing survey sites within the BBS bioregion were utilised by the JVMP. Figure 2 illustrates the locations of existing plot-based floristic survey sites in the BBS bioregion.

3.1.3 Abiotic variables used in gap analysis

Abiotic variables were used by the gap analysis tool to define the environmental and geographic space from which to select the survey sites. All abiotic variables used by the JVMP were required to have full coverage of the BBS bioregion. The following abiotic data layers were used in defining the environmental space.

- Mean annual temperature (continuous variable).
- Mean annual rainfall (continuous variable).
- DLWC draft protosoils layer or Murray Darling Basin Commission (MDBC) soil landscape mapping 1:250 000 (categorical variable).
- CTI (compound topographic index) wetness index (continuous variable).
- Geographic distance (continuous variable - distance between two points in the GIS).

Figures 3 to 5 illustrate the temperature, rainfall and wetness indices utilised by the JVMP.

Climatic variables were derived by NPWS GIS Unit (Hurstville), using ANUCLIM climatic modelling software (Houlder *et al*., 1999). Soil properties were derived from the Murray Darling Basin Commission, Basin in a Box, 25m GIS dataset (MDBC, 2000) and later, when available, the DLWC BBS bioregion Soil Landscape Reconnaissance Mapping (RACAC, 2002).Wetness indices were derived by the NPWS WRA Unit using a 25m digital elevation model (DEM) supplied by NPWS GIS Unit (Hurstville). Geographic distance was derived by the NPWS GIS Research and Development unit (Armidale).

3.1.4 Survey site selection

Candidate sites

Candidate sites are a suite of sites from which the survey sites are selected. Due to the limits of computer hardware and software architecture it is preferable to select a sample of candidate sites from within the surveyable domain (ie, the BBS bioregion) and to then select the survey sites from the set of candidate sites, using the gap analysis tool (Section 3.1.1, and Appendix 11). Candidate sites are randomly selected at the start of the process by simple constrained random number generator for easting and northing values. Figure 6 provides and example of 20 000 candidate sites locations selected for the JVMP.

Survey site selection

Survey sites were selected from the suite of candidate sites using the gap analysis tool. Existing sites were highlighted by the tool to minimise duplication. Survey effort was targeted at those candidate sites, within the environmental and geographical space, accorded the highest level of priority for survey. Survey site priority was determined by the uniqueness of the combination of environmental variables and the amount of previous survey effort within such environmental space.

3.1.5 Masking of target areas

The JVMP used a number of masks during the gap analysis. Masks constrained the survey effort to parts of the surveyable domain of most interest.

Masks were originally omitted from site selection so that the most important sites across the bioregion could be selected, without reference to tenure. However, the high cost of survey required that masks be applied because many of the survey sites were falling within areas of little or no native vegetation.

- The initial gap analysis was constrained to areas of woody vegetation in an effort to commence survey work early.
- Gap analyses 2 and 3 were unmasked with sites selected across all tenures.
- Gap analysis 4 had a preliminary woody vegetation mask applied to constrain the selection of sites to areas of identifiable native vegetation.
- Gap analysis 5 utilised an updated woody vegetation mask.
- Gap analysis 6 was conducted without a mask in an effort to capture information in areas of predominantly very open woodland and/or grassland.

Figure 7 depicts the initial woody vegetation mask as used in gap analysis 4.

FIGURE 2. Locations of existing plot-based floristic survey sites in the BBS Bioregion

Includes BBS Bioregion Stage 1 Survey Plots (Figure 1) And Plots From Other Sources

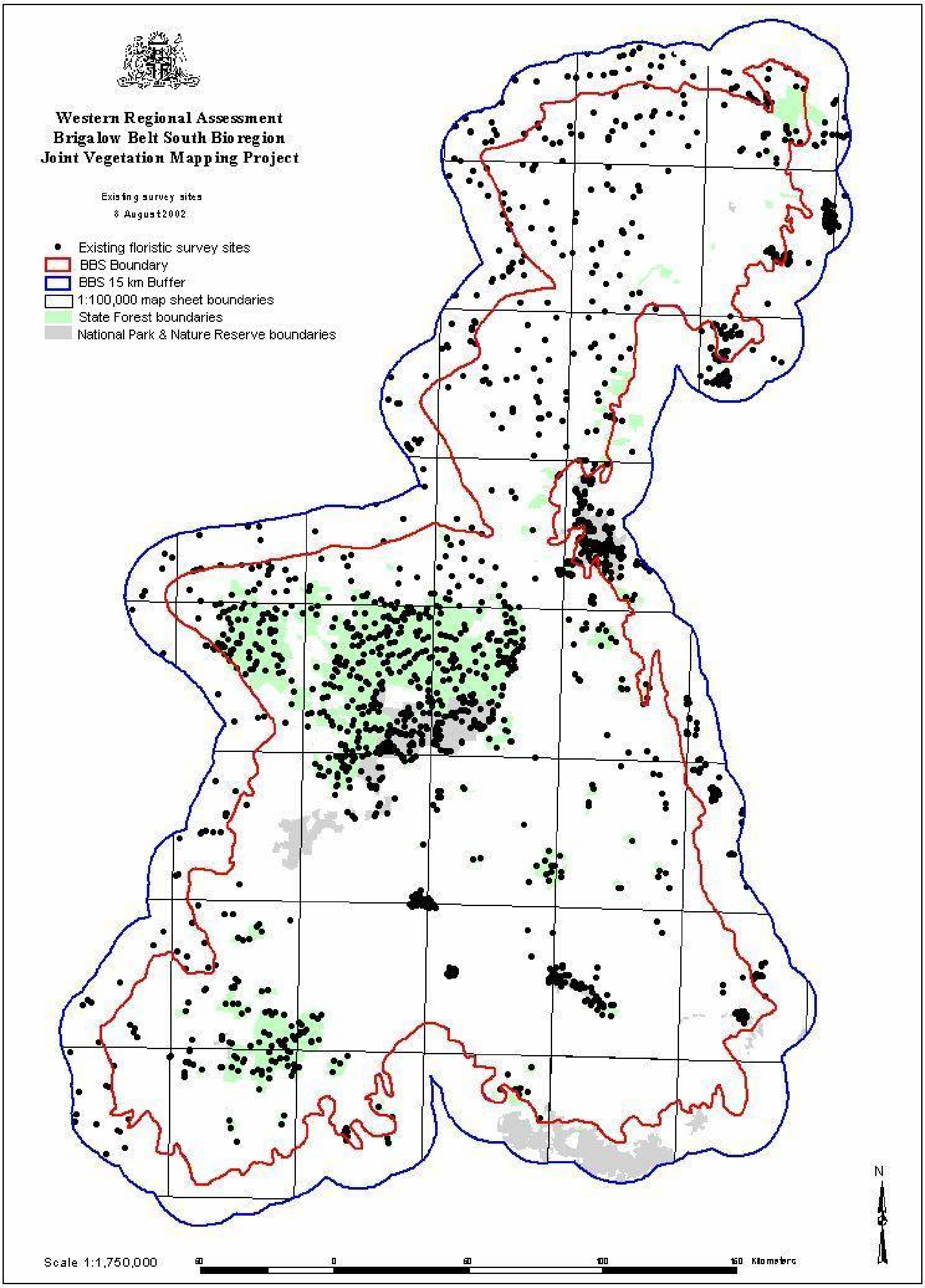


Figure 3: Wetness index as used in gap analysis

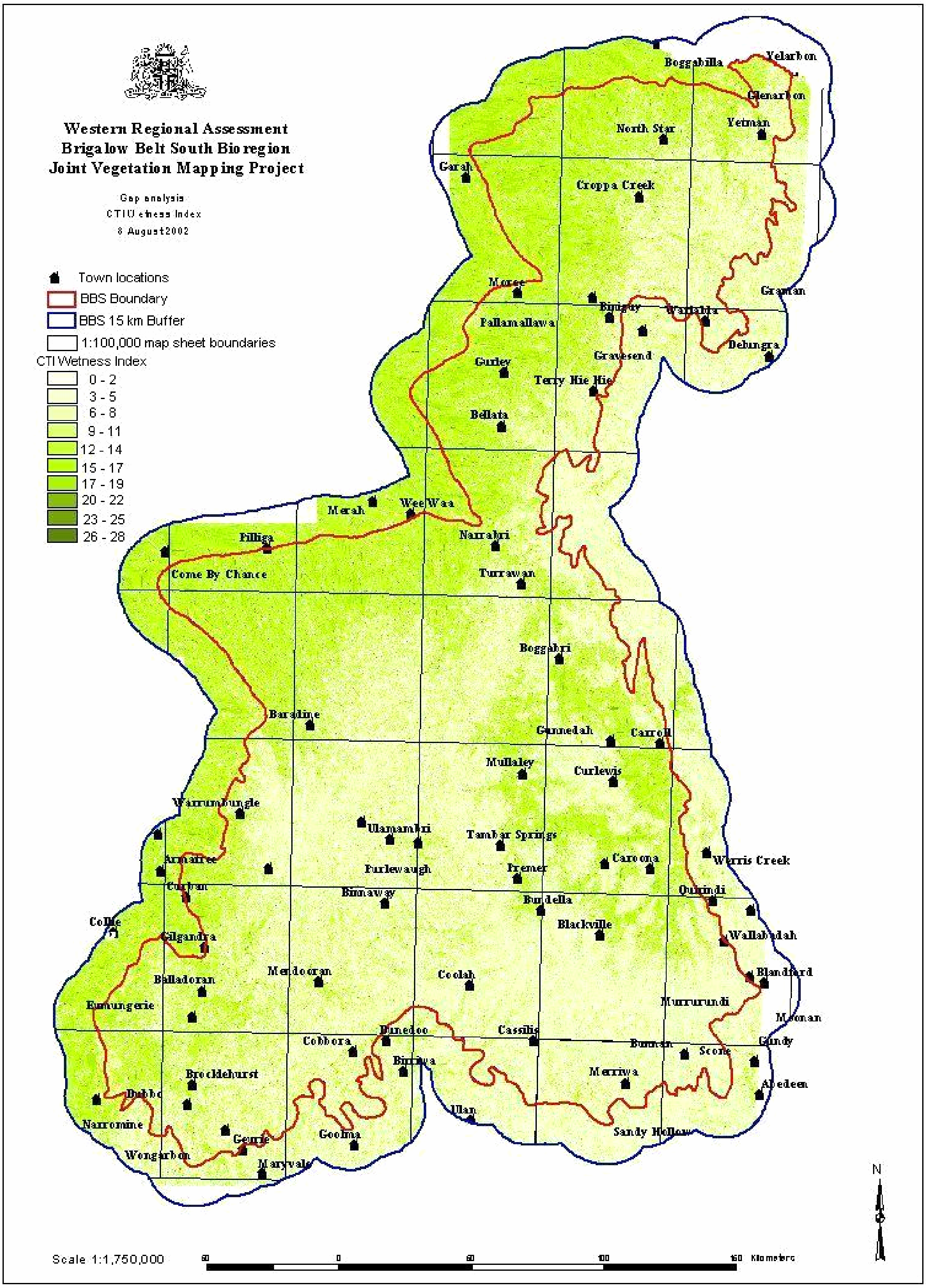


Figure 4: Mean annual temperature as used in gap analysis

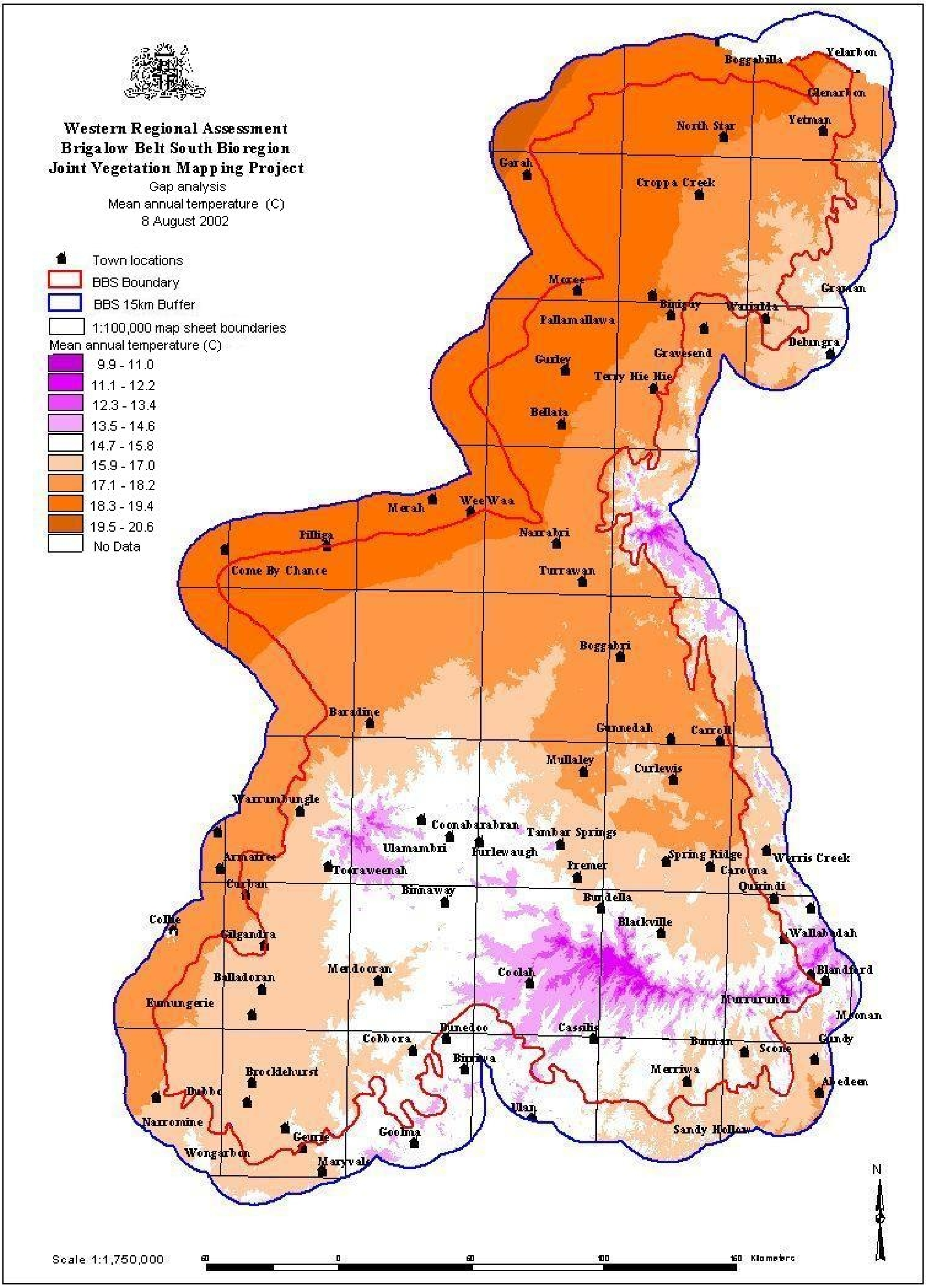


Figure 5: Mean annual rainfall as used in gap analysis

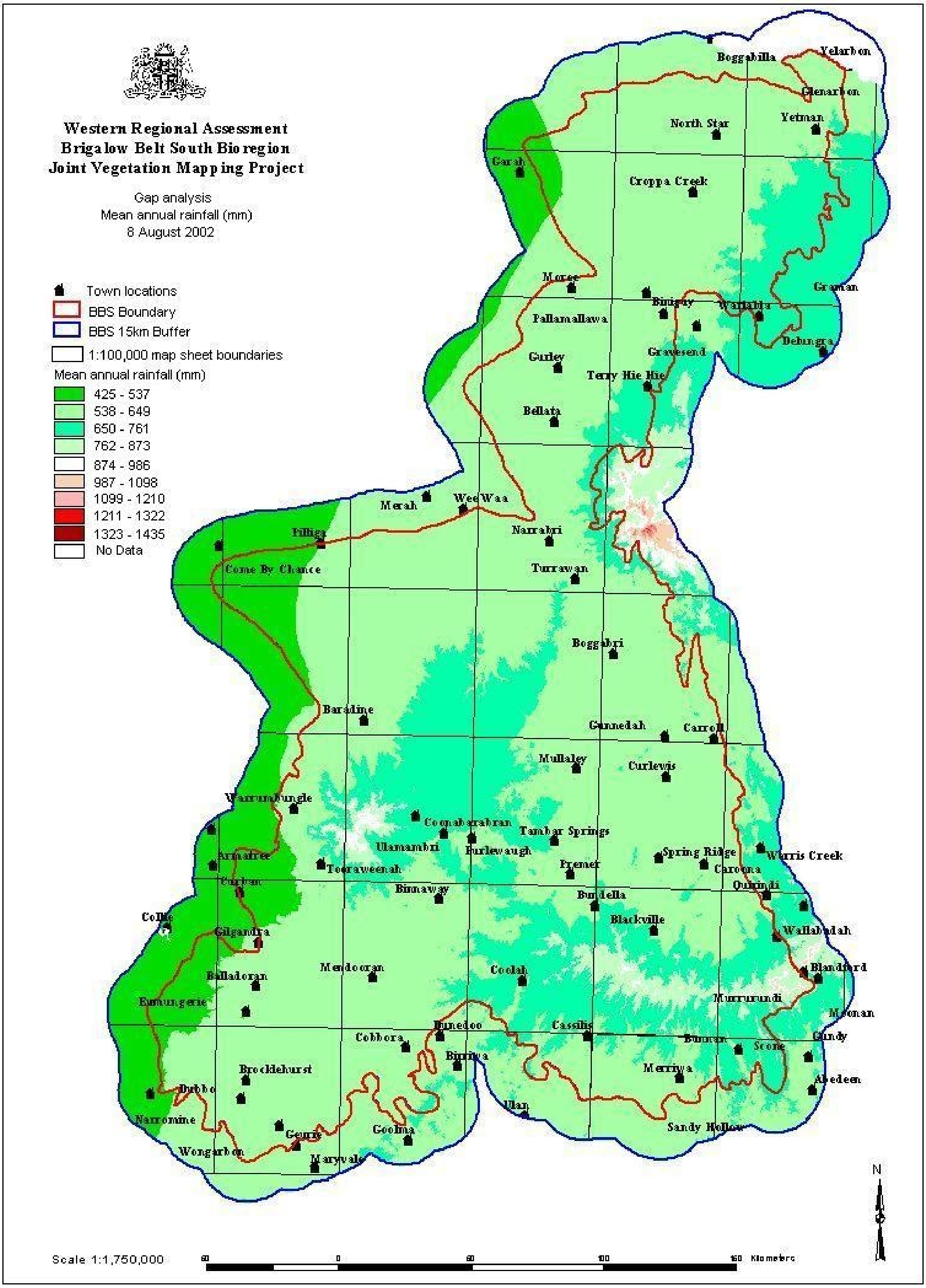


Figure 6: Candidate site locations for gap analysis

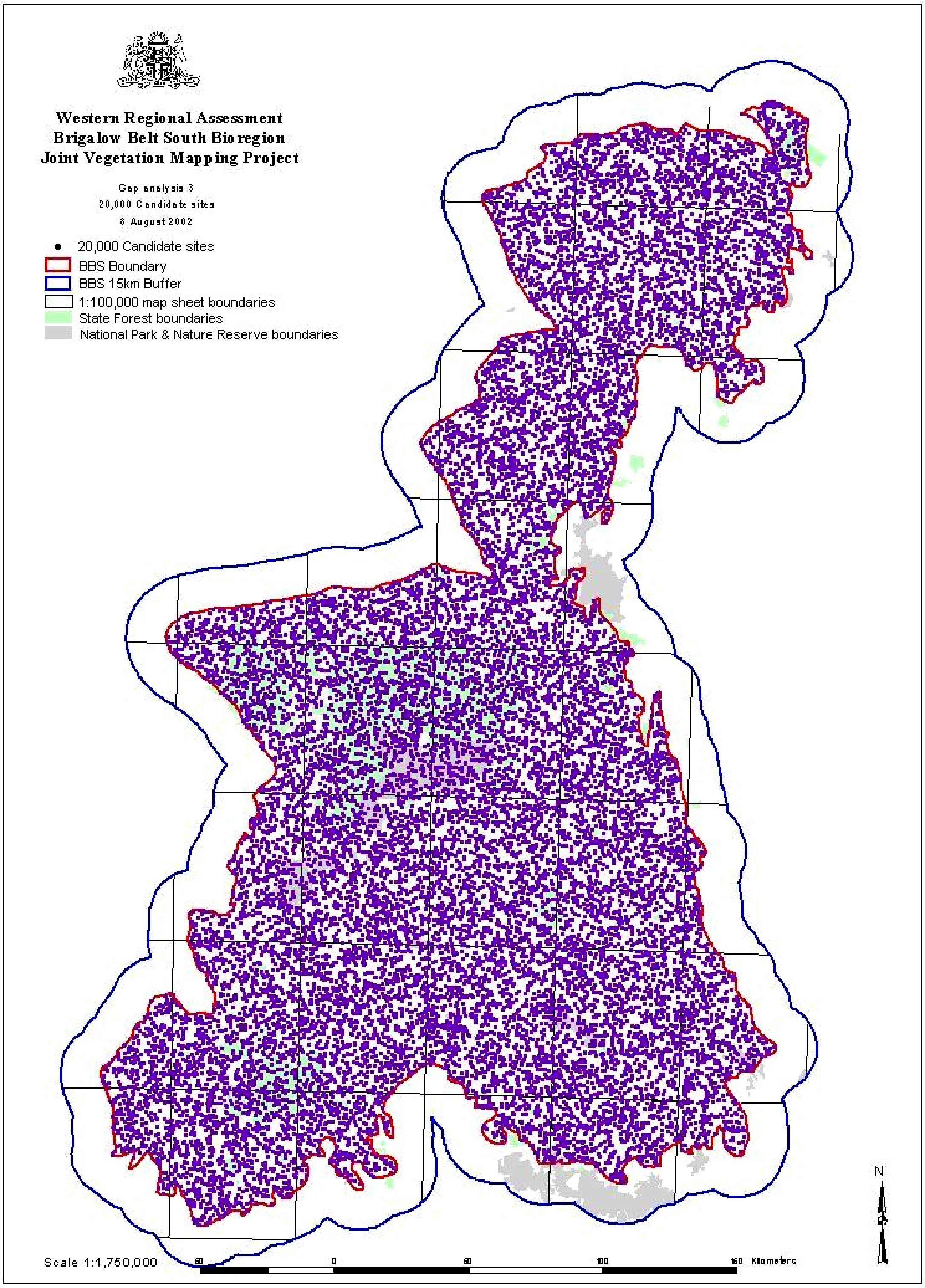
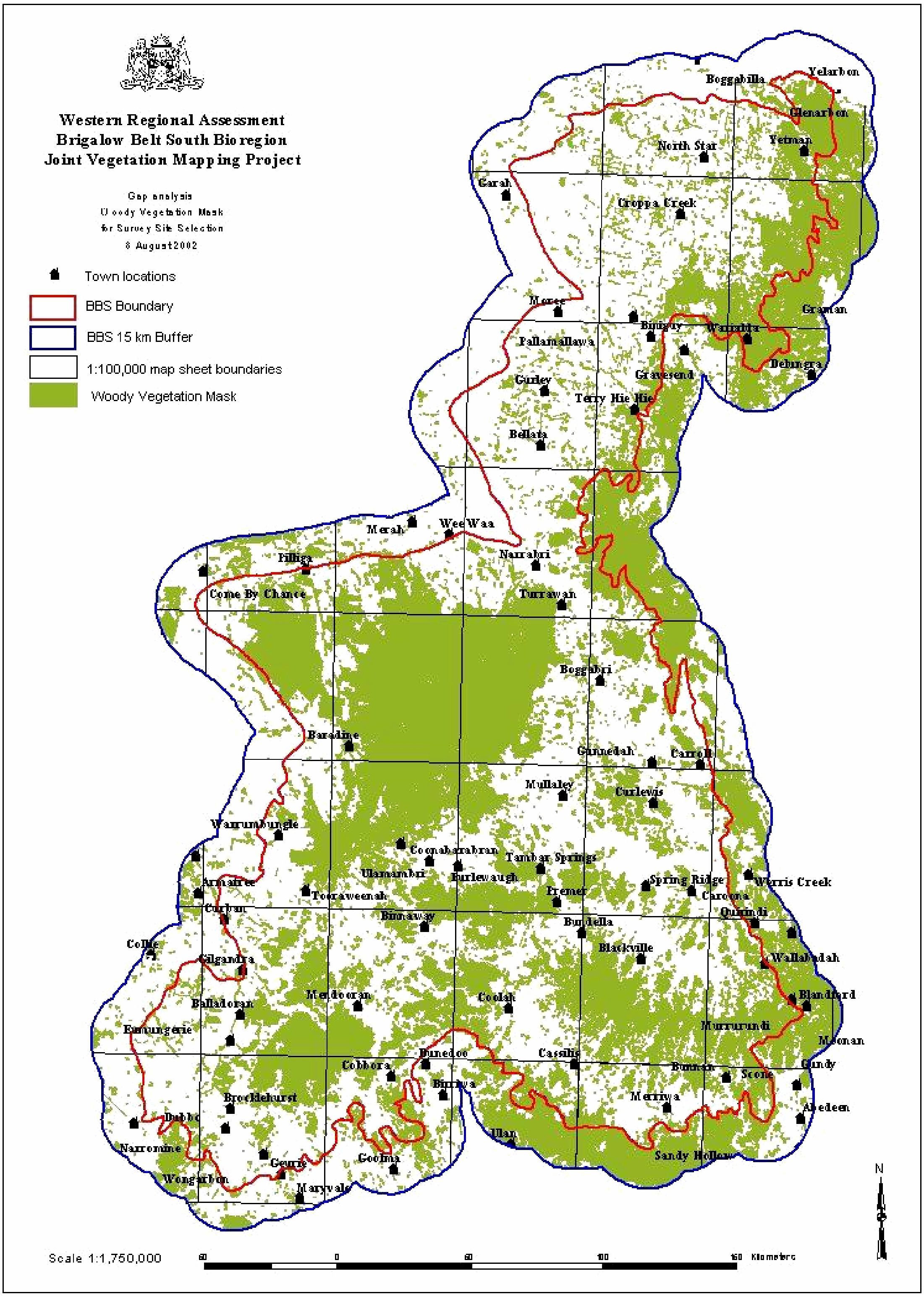


Figure 7: Woody Vegetation Mask used in gap analysis 4



3.2 FLORISTIC SURVEY

3.2.1 Survey standards

The floristic survey was designed to meet the technical standards set down in the DLWC *Guidelines for mapping native vegetation* (Sivertsen and Smith, 2001). Survey proformas designed by DLWC were utilised in the initial stage of the project. However these did not meet the requirements of all partner agencies and a modified vegetation survey proforma was developed for use in the JVMP. This did not impact upon the utility of the survey work carried out early in the JVMP as each survey was treated as distinct for the purposes of data storage and analysis.

3.2.2 Survey timeframe

The initial survey commenced during the summer of 2000-2001 to take advantage of favourable survey conditions. DLWC carried out further sampling for the NVMP in the north of the bioregion from this time. Full floristic JVMP survey commenced after delivery of the first gap analysis results.

Floristic survey calibration field days were held in September 2001 at Baradine. Survey proformas were tested in the field and changes made for clarity and to meet partner agency requirements (refer Appendix 1 for field proforma details). Agency botanists and contract botanists resolved differences in interpretation of the methods and a level of consistency was achieved.

3.2.3 Survey database

Vegetation survey data collected included floristic, physiographic and structural information. The Vegetation Survey Database (VSD) developed by the NPWS was used to store and access data that met the NPWS floristic survey criteria. Additional data routinely collected for the DLWC Native Vegetation Mapping Program (NVMP) was stored in a specifically developed relational database known as YOWIE. This allowed the additional structural data to be captured. Both the VSD and YOWIE are based on the MS Access 97 platform. These relational databases can maintain links between tables and allow queries, functions and macros to be utilised in the interrogation and interpretation of data. Refer to section 3.6 for additional information on the databases.

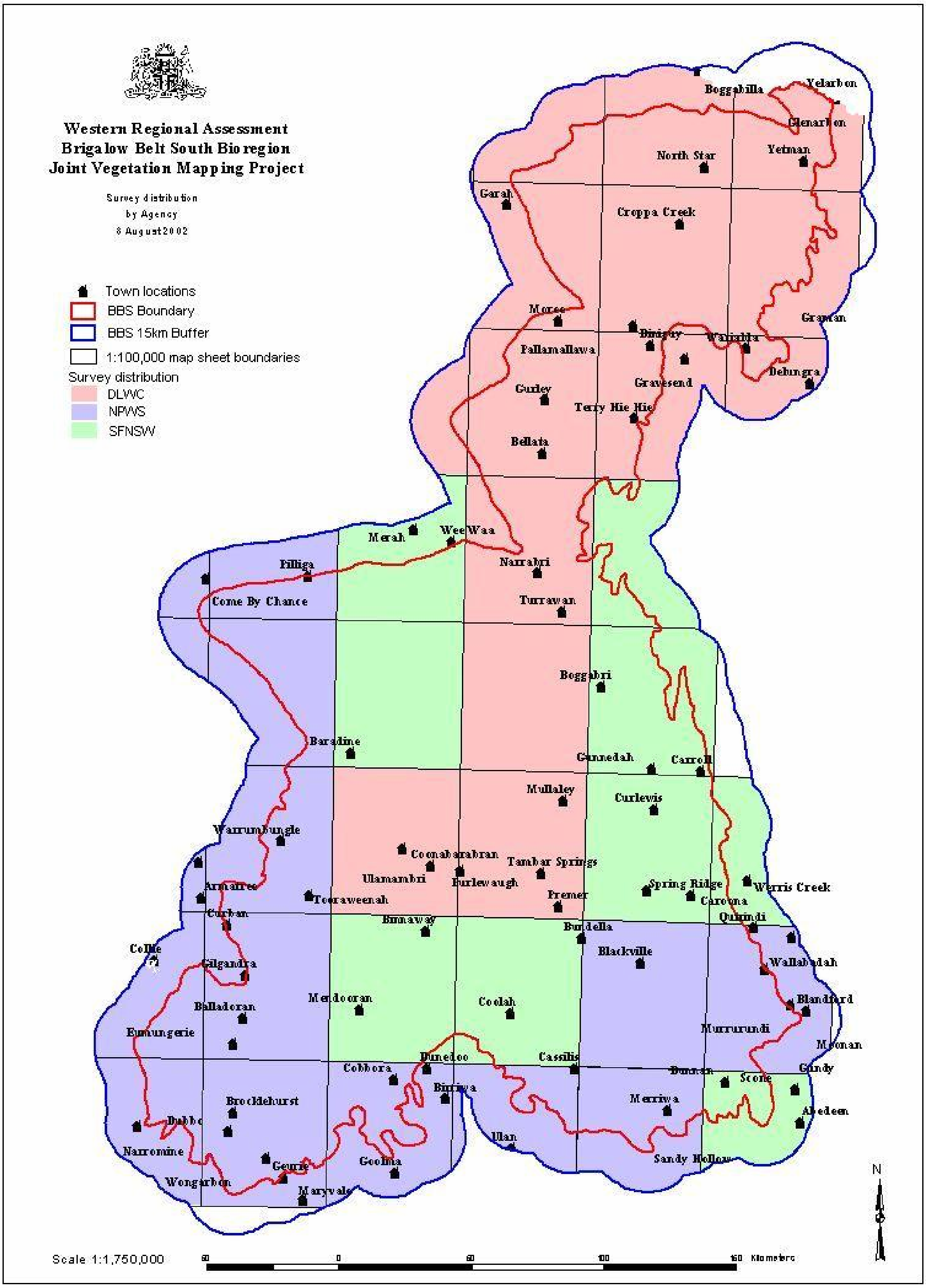
3.2.4 Survey effort distribution

Protocols were established for the distribution of survey sites to the partner agencies and for the delivery of completed survey forms on a fortnightly basis. Figure 8 illustrates the distribution of survey effort by agency. Survey sites were provided by NPWS as outputs from the gap analysis. Botanists met regularly to confer on species identification. Where ambiguity remained, vegetative samples were forwarded to the Royal Botanic Gardens (RBG) for formal identification.

3.2.5 Survey completion

Floristic surveys ceased in the north of the bioregion in early February 2002 due to drought conditions impacting upon species number and making identification of remaining plants to species level unreliable. All JVMP floristic survey within the BBS bioregion had ceased by the 31 March 2002 so that data entry timetables could be met.

Figure 8: Distribution of survey effort by agency



3.3 AERIAL PHOTOGRAPHY INTERPRETATION

Targeted aerial photography interpretation (API) was carried out within the BBS bioregion. Targeted API was essential to the project as it provided overstorey pattern information which could be used to constrain the vegetation model (refer section 3.12.4 for details on model constraints).

Full API coverage of the BBS bioregion could not be achieved due to budget and time constraints. The JVMP TWG decided to undertake targeted API complementary to the NVMP API. The NPWS WRA Unit, Dubbo supervised this work.

3.3.1 Data audit of API coverage

An audit of existing API datasets and other ongoing API showed gaps in the coverage of API across the bioregion. The audit involved a review of metadata and accompanying reports. It included consultation with botanists to ascertain the range of available data sets and their respective currency, custodianship, coverage and reliability. Discussions were held with DLWC to resolve the extent of the NVMP API program and its timeframe.

Once the extent of existing and proposed API was known it was possible to identify targeted API priority areas. Figure 9 illustrates the coverage of recent API, scheduled NVMP and Targeted API at the commencement of the project.

3.3.2 Data audit of aerial photographs

Aerial photography was sourced from the relevant agencies or from Land Information Centre. Photographs were 1:50 000 scale or 1:25 000 scale and dated 2000-2001.

3.3.3 API map attributes and mapping pathway

The development of the API mapping pathway was based on the specific requirements of the project brief, namely: to provide full floristic and structural data that met the NVMP *Guidelines for mapping native vegetation*. The JVMP TWG assessed the three API mapping pathways being used within the BBS bioregion by DLWC and NPWS for NVMP API survey. The TWG determined that the NPWS (Northern Directorate) API pathway was the most appropriate for fulfilling the requirements of the JVMP in the timeframe available. The API pathway included vegetation cover with more than 10% canopy cover.

The mapping pathway for targeted API within the BBS bioregion was based firstly on vegetation cover, then overstorey floristics, juvenile canopy cover (growth stage), understorey type, canopy height, disturbance and land use. Thresholds applied to canopy cover (10% ccp) and minimum polygon size (10ha), with exception for special features (2ha). The JVMP targeted API mapping pathway is illustrated in Appendix 2.

3.3.4 Targeted API map sheets

A number of map sheets were already planned for completion under the NVMP API program. After considering the requirements for additional API within the BBS bioregion, the JVMP TWG decided to target the Cobbora, Gulgong, Merriwa, Blackville and Murrurundi 1:100 000 map sheets in the south of the BBS bioregion and the Bingara, Yetman and Yallaroi map sheets in the north. API programs underway within and adjoining the BBS bioregion are illustrated in Figure 9.

3.3.5 Data requirements

Data requirements to be delivered by the API contractors consisted of the following components.

- Photo preparation (eg, markup photos for study area boundary, affix overlays, mark fiducials).
- Aerial photograph interpretation (the stereoscopic interpretation of vegetation structure, floristics, disturbance and land use) with all linework containing a unique polygon identification and edge matching for each map sheet.

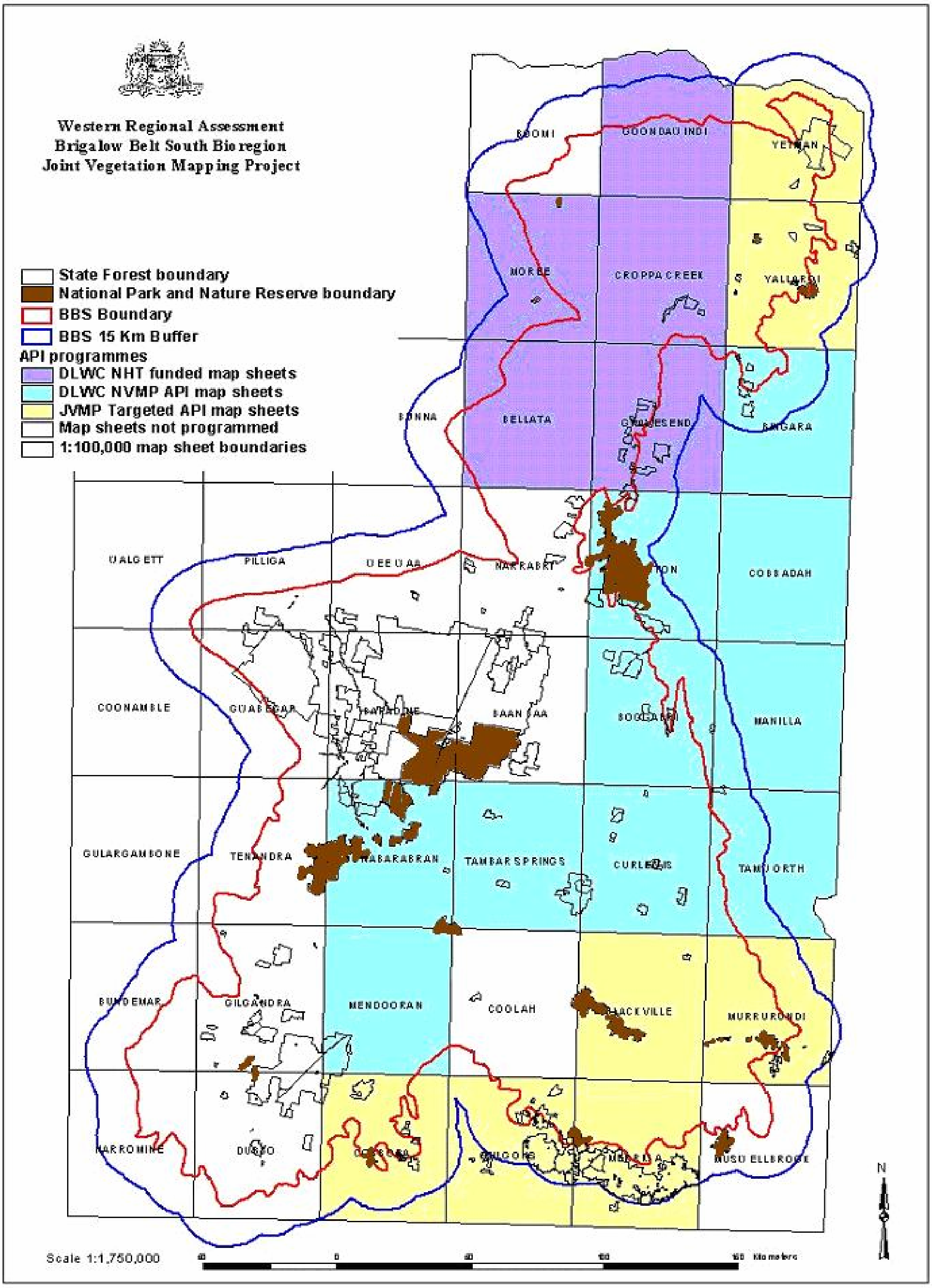
- Ground truthing (including completion of survey proformas in either digital or hardcopy formats and a brief contextual write-up of the areas where ground truthing was undertaken).
- Compilation of results (summary of API results, checking all linework and polygons).
- Providing all the above components within the specified time frames.

3.3.6 Consistency

Primary controls for consistency in API mapping included:

- contracting experienced API interpreters with extensive knowledge of the vegetation types occurring in the bioregion; and
- conforming to the API mapping pathway determined by the TWG.

Figure 9: Priority API programs within and adjoining the BBS bioregion



API interpreters, botanists, project managers and GIS support staff attended a calibration workshop. The participants agreed on key project parameters, including:

- standardised pathways for code strings, methods for labelling polygons with unique identifiers, reference codes for areas with poor floristic interpretability, and standards for maintaining floristic code running sheets and documenting pattern characteristics such as colour, tone, texture, shadow and density;
- realistic timeframes for the delivery of mapping;
- clear guidelines for access to private land consistent with principles outlined by RACAC. The intent of these principles was to ensure that any ground surveys carried out on private land was at the landholder’s discretion;
- clear guidelines for setting ground control reference points and for using geographical Positioning Systems (GPS);
- decision to use recent, high quality aerial photographs available (LIC dated 2000 to 2001); and
- establishment and maintenance of good communication links between all participants in the API process.

3.4 NVMP API

An objective of the JVMP was to produce six 1:100 000 map sheets to NVMP standards using the DLWC *Guidelines for mapping native vegetation, Version 2.1* (Sivertsen and Smith, 2001). As part of this project, the following map sheets were scheduled for API mapping to DLWC NVMP standards:

- Boggabri
- Gravesend
- Curlewis
- Tambar Springs
- Coonabarabran
- Mendooran (However, due to resourcing problems this map sheet was subject to targeted API of the woody vegetation only).

These six map sheets, together with the targeted API, formed the core of the composite API vegetation data layer. Figure 9 illustrates the priority NVMP API areas within and adjoining the BBS bioregion.

The Mendooran 1:100 000 map sheet was considered to be a core map sheet for inclusion in the JVMP and was therefore included in the targeted API program and contracted out to experienced interpreters.

The contractors utilised traditional methods for the field work component and linework development, following the API mapping pathway as described in section 3.3.3 and Appendix 2. A new approach to data capture was adopted whereby the interpreters directly captured their linework and attributes into a Geographical Information System (GIS). This process provided numerous advantages including substantial cost savings over traditional linework capture methods, and a reduction in errors due to duplicated effort when either digitising or scanning API overlays.

Linework was digitised by the interpreter in a hybrid data capture system which still required the use of a stereoscope and hard copy aerial photography. The photographs were interpreted directly when seated at the computer and then digitised directly on to an orthorectified satellite image displayed on the computer screen. This method also provides a greater level of security for original linework as the linework can be saved to a read only format CD-Rom. This medium is easy to store and copies of the original work can be provided as required.

3.5 COMPOSITE API LAYER

Approximately 30 API and vegetation datasets were used to produce a composite API layer (Table 1). The layer comprised existing datasets and those produced by the JVMP and NVMP (see figure 10).

The composite API vegetation layer was used as a constraint or conditioner upon the Generalised Dissimilarity Model. API compliments floristic survey effort by providing information about the structural patterns of the modelled vegetation groups not available through floristic sampling or mathematical modelling.

Digital API data sets were sourced and ranked according to currency, resolution and quality. Older data sets were assumed to have less utility than recently completed API data sets.

Resolution referred to a combination of scale of photography and data capture, as well as the level of identification of vegetation (for example, ranging from species level down to broad vegetation group) or within the data set.

Quality of the data referred to a combination of currency and resolution including the level of detail captured. For example, some data sets provided a complete suite of attributes that could be used to enhance the modelling, while others provided only a broad common descriptor and revealed little about the underlying nature of the vegetation being observed.

Due to the differences in currency, resolution and quality the composite API layer was required to use the lowest common denominator as the basis for its construction, ie overstorey species identification.

TABLE 1: COMPOSITE API DATA LAYERS, BRIEF DESCRIPTIONS AND CUSTODIANS

Dataset name	Description	Custodian
Bingarra-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
Binnaway-npws	Binnaway NR vegetation mapping	NPWS
Blackville-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
Gravesend-nvmp	Native vegetation mapping program	DLWC
WRA Stg 1 API	Targeted API — woody vegetation within State forests and national parks	NPWS
NWB 2000	Wheatbelt mapping 2000 update	NPWS
Yetman-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
Yallaroi-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
sftype-sfnsw	State Forests forest typing	SFNSW
Coonabarabran-nvmp	Native vegetation mapping program	DLWC
Curlewis-nvmp	Native vegetation mapping program	DLWC, NPWS, PlanningNSW, SFNSW
Tambar Springs-nvmp	Native vegetation mapping program	DLWC
Bugaldie-nvmp	Native vegetation mapping program	DLWC
Boggabri-jvmp	Native vegetation mapping program	DLWC, NPWS, PlanningNSW, SFNSW
Moree shire-dlwc	Moree shire mapping	DLWC
Mullalley-nvmp	Native vegetation mapping program	DLWC
Cobbora-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW

Mendooran-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
Murrurundi-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
Gulgong-jvmp	Targeted API — woody vegetation	DLWC, NPWS, PlanningNSW, SFNSW
Mt Kaputar-npws	Mt Kaputar NP vegetation mapping	NPWS
East Walgett-dlwc	East Walgett shire mapping	DLWC
Pilnatres-npws	Pilliga nature reserve mapping	NPWS
grnpmgnr-npws	Goulburn river NP and Munghorn Gap NR vegetation mapping	NPWS
Weetalibah-npws	Weetalibah NP vegetation mapping	NPWS
Warrumbungles-npws	Warrumbungles NP vegetation mapping	NPWS
Towarri-npws	Towarri NR vegetation mapping	NPWS
NWB 1994	Wheatbelt mapping 1994 update	NPWS
CRAFTI-CRA	CRA LNE data	NPWS, PlanningNSW, SFNSW
BBS bioregion Landuse mapping	DLWC landuse dataset	DLWC
MDBC-M305	MDBC-M305 floristics	MDBC
Rusden-M305	MDBC-M305 floristics	MDBC

3.6 DATA ENTRY

3.6.1 The NPWS Vegetation Survey Database (VSD)

The NPWS Vegetation Survey Database (NPWS, 2002) consists of four main components:

- the General Section contains locality information for each site and information that characterises site accessibility (for example, land tenure);
- the Floristics Section records data on vegetation community structure and floristics (the individual species found within the survey area);
- the Physical Section records environmental information such as terrain, climate, lithology, soils and hydrology;
- the Disturbance Section records any physical disturbance at the site, including fire, grazing, logging and any other perturbation.

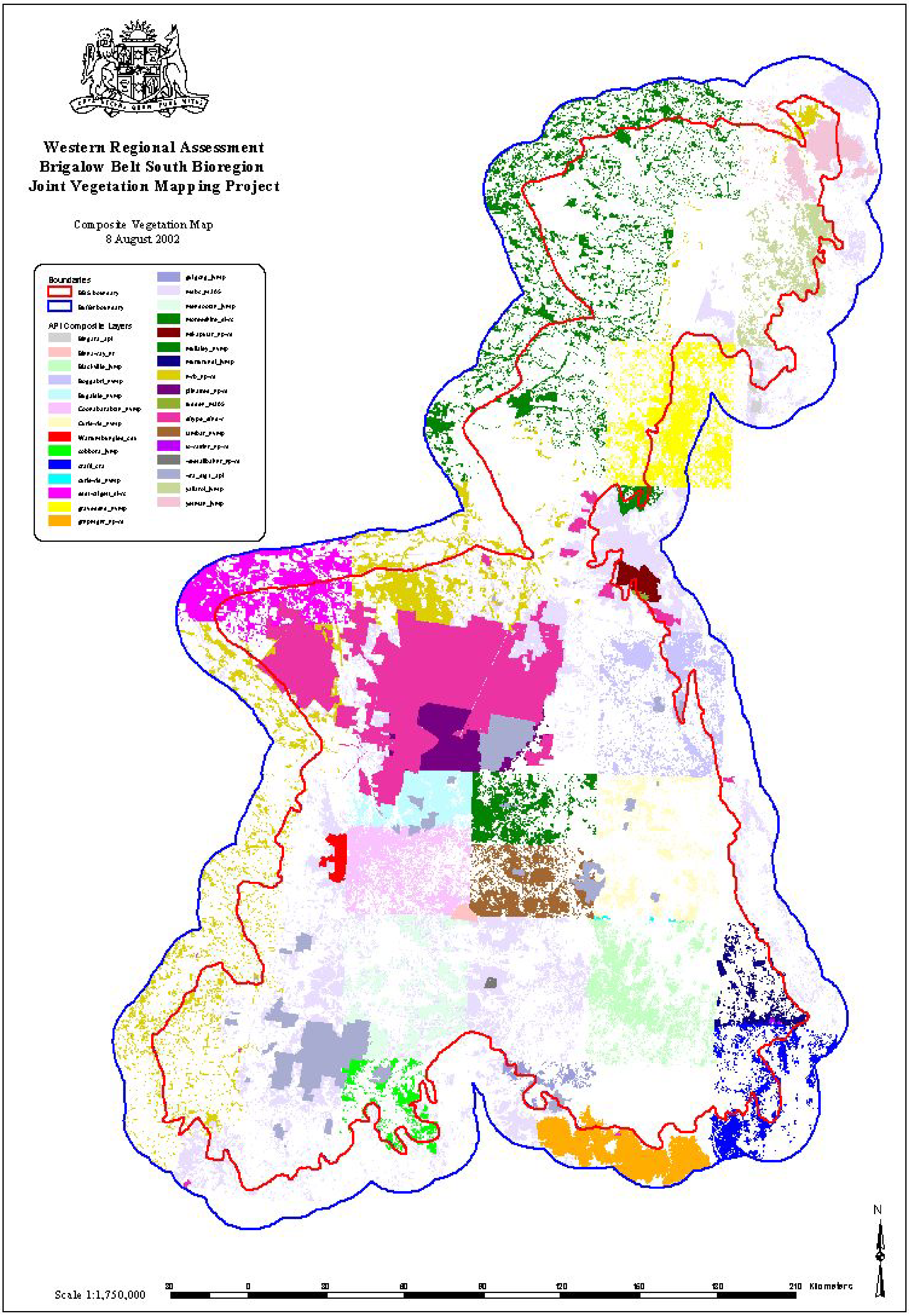
3.6.2 The YOWIE database

The new proformas designed for the JVMP vegetation surveys included a number of data fields that could not be entered into the VSD, such as tree diameter and certain soil characteristics. This non-standard data was entered into a new MS Access database (YOWIE) designed by the NPWS WRA Unit and extensively upgraded by DLWC.

3.6.3 The combined JVMP data set

The JVMP utilised an initial combined data set of 3166 floristic records. The data set contained the existing floristic survey data as determined by the data audit (refer section 3.1.2), the JVMP floristic survey sites and the NVMP floristic survey sites which fell within the BBS bioregion boundary.

Figure 10: API layers utilised in composite layer production



3.7 DATA PREPROCESSING

3.7.1 Recoding and standardisation of cover-abundance scores

All data used for this project included a quantitative estimate of the amount of each species in each survey site, usually either as a combined cover-abundance score or as separate estimates of cover and abundance. In addition to variations in whether cover and abundance were recorded separately or as part of a combined measure, surveys differed in the number and size of classes to which estimates were allocated. The definition of cover was not specified for most surveys. Most observers assess the combined perpendicular projection of all aerial parts of all individuals of the species being assessed, usually referred to as canopy cover. However, some surveys use crown cover (perpendicular projection produced by assuming opaque crowns) or both canopy cover and crown cover. Cover and abundance estimates are made visually in the field and, even with a consistent method of scoring, may vary considerably among observers and for a single observer at different times.

For consistency, the various scoring and assessment methods were converted as far as possible to a common basis. For most survey sites, the scores were based on, or could be easily converted to, a modified Braun-Blanquet scale, such as:

- 1 = uncommon or few individuals and up to 5% cover;
- 2 = any number of individuals and up to 5% cover;
- 3 = 6-25% cover;
- 4 = 25-50% cover;
- 5 = 51-75% cover;
- 6 = 76-100% cover.

This was adopted as the basic standard for analysis and all survey data was recoded to match these codes as closely as possible. Where it was not specified or was otherwise unknown, cover was assumed to refer to canopy cover, but high values for some surveys suggest that crown cover may have been used. The six types of data conversions applied are described in Table 2 and conversions applied to each of the survey data sets are listed in Table 3.

3.8 FLORISTIC RECODING

Floristic data varied in taxonomic and nomenclatural recency, degree of taxonomic resolution and accuracy. To overcome this variation all data were validated and where necessary recoded. The data were recoded using the steps described below, to maximise consistency while minimising loss of taxonomic resolution.

- Update nomenclature to a consistent standard where no ambiguity existed. This included nomenclatural changes without taxonomic division and taxonomic changes that could be unambiguously assigned to all records, for example those specific to geographical location. Most of this step was done with automated routines and a standard taxonomic and nomenclatural reference list (the CAPS list maintained by NSW NPWS).
- Aggregate all infraspecific taxa at the specific level. Although this resulted in some loss of taxonomic information, it was preferable to losing the substantial proportion of survey data for which infraspecific taxa were either not recorded, or not consistently recorded.
- Change or omit taxa considered likely to have been misidentified, but for which verification was not possible or practical.
- Aggregate taxa for which consistent and reliable field determination was considered difficult or unlikely, and for which recent taxonomic changes rendered older records ambiguous.

- Remove records for taxa recorded only to genus or higher taxonomic levels, or not identified to any clear taxon.

TABLE 2: CONVERSION OR RECORDING OF COVER-ABUNDANCE SCORES

Conversion type number	Assessment method used for original data	Description of conversion rules
	Cover-abundance code 1-6 as defined by above data analysis standard; definition of cover unknown or specified as canopy cover.	No conversion.
0.	Cover-abundance code 1-6 as: 1 = uncommon; 2 = common and cover up to 5%; 3 = 6–20% cover; 4 = 21-50% cover; 5 = 51-75% cover; 6 = 76-100% cover.	Cover codes accepted as equivalent, ignoring different class limits for classes 3 and 4. This difference was regarded as minor in the context of variation among observers and was thus assumed to have negligible effect on the results of analysis.
1.	JVMP standard assessment method as described under field survey. Cover and abundance recorded separately. Cover usually estimated to nearest 5% class or as <1% if applicable, although sometimes to nearest 1%. Abundance estimated as number of individuals in one of six classes.	If canopy cover is >5%, allocate to matching broader class (cover-abundance codes 3-6); if cover <=5%, allocate to cover-abundance =1 for abundance classes 1-2 and cover-abundance =2 for abundance classes 3-6.
2.	As for type 1, but cover known to be assessed only and specifically as crown cover.	As for 1, but initial cover multiplied by 0.6 prior to allocation to cover classes, as an average estimate of the relationship between crown cover and canopy cover for the range of species recorded.
3.	As for 2, but separate recording of crown cover and canopy cover inconsistent.	As for 1 where canopy cover was recorded; otherwise as for 2 if only crown cover was recorded.
4.	Cover-abundance on 7-point scale, as: 1 = one individual; 2 = few individuals and <5% cover; 3 = numerous individuals and <5% cover; 4 = 5-25% cover; 5 = 25-50% cover; 6 = 50-75% cover; 7 = 75-100% cover; Cover not defined.	Allocate classes 1 and 2 to 1, 3 to 2, 4 to 3, 5 to 4, 6 to 5, 7 to 6.
5.	As for 4, but for trees (woody plants taller than 2 metres) only number of stems was recorded.	As for 4, but number of stems for trees was converted as: 1-2 stems allocated to cover class 2; 3-10 stems to class 3; 11-30 stems to class 4; 31-60 stems to class 5; >60 stems to class 6.

TABLE 3. DATA CONVERSIONS

Survey ID	Cover-abundance Conversion Type
ARAKoola	Not modified
Astro	Not modified
BBS	2
BBSCOMM	3
BBSINV	1
BIN_NR_97	0
COOLAH_93	Not modified
DAP	0
DRP_2000	1
DUBBO_99	Not modified
EWDRP	2
GOONOO_99	Not modified
jvmpbcvl	2
JVMPDB	1
JVMPEA	1
MAC	Not modified
MOREEGRASS	Not modified
MTKAP2000	Not modified
MTKAPA_97	0
NAMOI_95	Not modified
NAN	2
Narrom_99	Not modified
NBAFF	Not modified
NCPP	Not modified
NVMP-INV	1
NWB	4
OOLINE	5
PIL	4
PIL_NRa_99	Not modified
PIL_NRB_99	Not modified
PIL_SF_95	Not modified
PILL_b_99	Not modified
PILLC_99	Not modified
PILLIGA_99	Not modified
PLAINSF_99	Not modified
RM_JVMP	1
RMDRP	1
STH_KAP_98	Not modified
Towarri99	Not modified
TPJVMP	1
TSUTFS0001	0
WEET_NR_97	Not modified

.

Taxonomic and nomenclatural changes were made for taxa considered to have been misidentified and when verification was not possible or practical (table 4). Where consistent and reliable field determination was considered difficult or unlikely, and recent taxonomic changes rendered older records ambiguous, taxa were aggregated as described in table 5.

Further data preprocessing was as follows:

- all exotic taxa were removed prior to floristic analysis, on the basis that exotic species do not assist in characterising native vegetation groups;
- survey sites with very low numbers of species were excluded from analysis. The threshold was arbitrarily set at five, so that sites with five or fewer species were excluded.

TABLE 4: CHANGES FOR PROBABLE MISIDENTIFIED OR INCORRECTLY RECORDED SPECIES

Changed taxon	Reason for change
Acacia blakei to A. cheelii	Probable misidentification.
delete Acalypha capillipes	Probable misidentification.
delete Alectryon subcinereus	Probable misidentification.
Bossiaea foliosa to B. obcordata	Probable misidentification.
Bursaria longisepala to B. spinosa	Probable misidentification.
delete Commersonia fraseri	Probable misidentification.
Croton insularis to Adriana glabrata	Probable misidentification.
Dichelachne inaequiglumis to D. micrantha	Probable misidentification.
Dillwynia cinerascens to Pultenaea cinerascens	Probable data entry/recording error.
Dodonaea tenuifolia to D. falcata	Probable data entry/recording error.
Eucalyptus tereticornis to E. blakelyi	Probable misidentification.
delete Eucalyptus tindaliae	Probable misidentification or coding error.
Goodenia heterophylla to G. rotundifolia	Probable misidentification.
Lagenifera huegelii to L. gracilis	Probable misidentifaction.
Macrozamia pauli-guilielmi to M. concinna	Taxonomic change.
Marsdenia suaveolens to M. viridiflora	Probable misidentification.
Schoenus subaphyllus to S. kennyi	Probable misidentification.
Setaria paspalidioides to Paspalidium gracile	Probable misidentification of S. paspalidioides.
Wahlenbergia fluminalis to W. communis in Pilliga NR	Probable misidentification.

TABLE 5. TAXA AGGREGATED FOR ANALYSIS

Aggregate taxa	Reason for aggregation
Acacia uncinata s.l. complex	Recent taxonomic revision not readily applied to earlier records.
Acaena ovina, A. agnipila and A. echinata	Probable inconsistent field determinations.
Adiantum aethiopicum and A. atroviride	Recent taxonomic revision not readily applied to earlier records.
Cassytha glabella and C. pubescens	Probable inconsistent field determinations.
Chamaesyce sp. A and C. drummondii	Probable inconsistent field determinations.
Clematis glycinoides and C. aristata	Probable inconsistent field determinations of juvenile plants.
Dianella longifolia s.l.	Segregate taxa not consistently recognised.
Dianella revoluta s.l.	Segregate taxa not consistently recognised.
Dichelachne crinita and D. micrantha	Probable inconsistent field determinations.
Dichondra repens and D. species A	Probable inconsistent field determinations.
Einadia nutans and E. polygonoides	Probable inconsistent field determinations.
Enteropogon acicularis and E. ramosus	Probable inconsistent field determinations.
Eragrostis sororia and E. brownii	Probable inconsistent field determinations.
Laxmannia compacta and L. gracilis	Probable inconsistent field determinations.
Lepidium africanum and L. pseudohyssopifolium	Probable inconsistent field determinations.
Lepidosperma gunnii L. viscidum and L. laterale	Probable inconsistent field determinations.
Lomandra bracteata, L. cylindrica and L. filiformis	Probable inconsistent field determinations.
Marsdenia australis and M. viridiflora	Probable inconsistent field determinations of juvenile plants.
Melichrus sp. aff. erubescens and M. erubescens	Uncertain status and probable inconsistent field determinations.
Melichrus sp. aff. urceolatus and M. urceolatus	Uncertain status and probable inconsistent field determinations.
Olearia elliptica and O. sp. aff. elliptica	Segregate taxon not consistently recognised.
Pellaea falcata, P. paradoxa and P. calidirupium	Recent taxonomic revision not readily applied to earlier records.
Phyllanthus occidentalis and P. hirtellus	Recent taxonomic revision not readily applied to earlier records.
Picris hieracioides and P. angustifolia	Recent taxonomic revision not readily applied to earlier records.
Platsace linearifolia and P. sp. aff. linearifolia	Uncertain status.
Salsola kali and S. tragus	Recent taxonomic revision not readily applied to earlier records.
Sarcostemma australe and S. brunonianum	Recent taxonomic revision not readily applied to earlier records.
Stackhousia muricata and S. viminea	Probable inconsistent field determinations.
Verbena gaudichaudii and V. officinalis	Recent taxonomic segregation not readily applied to earlier records.
Vittadinia cervicularis and V. sulcata	Probable inconsistent field determinations.

3.9 DATA ANALYSIS

3.9.1 Floristic analysis

Floristic data were classified by grouping floristically similar sites using a numerical hierarchical agglomerative process. First, dissimilarity values were determined between all pairs of survey sites using the Bray and Curtis (Czekanowski) measure of association applied to the recoded but otherwise unstandardised cover-abundance data, using the ASO module of PATN (Belbin 1995). The Bray and Curtis measure has been shown to consistently perform well in recovering known relationships among test data (Faith, et al 1987). Sites were then grouped by applying a clustering algorithm with unweighted pair-group arithmetic averaging. The UPGMA routine in the FUSE module of PATN was used with a beta value of –0.1 (Belbin and McDonald 1993). Homogeneity analysis (Bedward et al 1992) was used to define a range of levels in the clustering hierarchy, represented by a dendrogram, from which to define floristic groups. The initial number of groups was chosen to be about twice the number indicated by the homogeneity analysis, so that finer scale detail could be examined. In particular, this gave the opportunity to aggregate fine-scale groups for which minor differences in floristic composition was considered to be due to artefacts of observer bias, disturbance, seasonal effects or a combination of these factors.

Following the initial clustering, a nearest neighbour check was conducted to identify potentially misclassified sites. Sites were regarded as potentially misclassified if three or more of the five most similar sites belonged to a different group.

Each of the reallocation rules listed in Table 6 were applied in turn to each set of sites meeting the criteria for misclassification. At each step, relationships among all sites were re-examined and the criteria and rules applied iteratively until no further reallocation was possible. Each step used less conservative criteria and rules than the preceding step. Reclassified sites were not further reclassified by a subsequent step, but could be reclassified by a subsequent iteration in one step. Criteria and rules were based on the five nearest neighbours (nnbs) to each site.

Following reallocations, the floristic composition, geographical distribution and environmental relationships of each of the resulting groups was then examined for the likelihood of observer, disturbance or seasonal artefacts. This was a subjective process, since sampling was not designed to formally test these influences. Where all or most sites in a group were from a single observer and the sites appeared to share very similar distribution and environmental features with an adjacent or closely similar group (judged from the group fusion distance in the hierarchy) for which there were different observers, the two groups were merged. Where a group comprised all or mostly highly disturbed sites with a large proportion of exotic species and appeared to occupy a similar distribution and habitat to another group, it was regarded as a probable artefact of past disturbance and merged with the most similar group.

Where a small group appeared to form primarily from having a common observer or high degree of disturbance, but relationships with other groups were unclear, plots were individually reallocated to the group to which the most similar plot belonged where the dissimilarity value was below 0.7. If dissimilarity of the most similar plot was above this threshold, the plot was left unallocated.

Further reallocations were made of plots for which the most similar neighbour was a different group, where the distance to this neighbour was <0.7 and where the next most similar neighbour was at least 0.05 more distant. Such plots were reallocated to the group of the most similar neighbour. Finally, the relationships of plots that appeared to be extreme geographical outliers were checked. These were reallocated based on a similar rule to the above. A plot was accepted as a real geographic outlier if the closest neighbour had dissimilarity <0.7 and there was a difference of at least 0.05 to the next closest neighbour. Otherwise, it was reallocated to the next most similar group to which it was geographically related, if the dissimilarity was below 0.7.

The final groups were checked by comparison with the results for the same number of groups produced using the non-hierarchical algorithm ALOC with the Bray and Curtis association measure (Belbin 1995). Both the weighted and unweighted mean within-group dissimilarity was used as a measure of the effectiveness of the final classification compared to raw outputs from the hierarchical and non-hierarchical clustering to the same number of groups.

TABLE 6: SUMMARY OF RECLASSIFICATION CRITERIA AND RULES

Misclassification criterion	Reallocation rule
All five nnbs in group other than site group.	Reallocate to group for which all five nnbs are in one group.
At least four nnbs in group other than site group AND most similar nnb in different group.	Reallocate to group for which the four most similar nnbs are in one group.
At least three nnbs in group other than site group AND two most similar nnbs in group other than site group.	Reallocate to group for which at least the three most similar nnbs are in one group.
At least three nnbs in group other than site group AND two most similar nnbs in group other than site group.	Reallocate to group for which the two most similar nnbs are in one group, providing at least one other nnb is in that group.
At least three nnbs in group other than site group AND the most similar nnb in group other than site group.	Reallocate to group to which the most similar nnb belongs, providing at least two other nnbs are in that group.

nnbs = “nearest neighbours”

3.10 DESCRIPTION OF FLORISTIC GROUPS

Profiles of floristic composition, vegetation structure and physical environment were prepared for each group based on summaries of data from the survey sites used for floristic analysis. Since vegetation structure was not consistently recorded for all sites, the profiles of vegetation structure were usually derived from subsets of sites. For floristic profiles, exotic species were included in the summaries, even though they were excluded from analysis, and aggregate native taxa were treated both as aggregates and as the originally recorded segregates.

Diagnostic species for each floristic group were defined by comparing the frequency of each species within each group to the overall frequency for all survey sites. Diagnostic species for each group were ranked using a binomial distribution, with the overall frequency for each species used as an estimate of the expected binomial probability. The two-tailed probability of obtaining a frequency at least as extreme as that observed for each species in each group was then used to rank the diagnostic value, with greater than expected frequency indicating positive diagnostic value and lower than expected frequency indicating negative diagnostic value. For descriptive purposes, species with high frequency or high median cover were included and ranked, even if not strongly diagnostic. In addition, all tree species with greater than five percent cover in any site were included in descriptions, regardless of frequency.

Vegetation structure and physical environment were characterised by calculating summary statistics (median, mean, percentiles and range) for each factor or component.

3.11 ABIOTIC VARIABLES USED IN THE MODELS

A suite of abiotic variables was utilised in the development of the JVMP vegetation models. Abiotic variables included climatic data (for example, rainfall and temperature), edaphic variables (for example, soils and soils attributes such as fertility), physiographic variables (for example, digital elevation models) and other variables. Only abiotic variables with full coverage of the BBS bioregion were used by the JVMP.

Abiotic variables were useful when establishing relationships with the survey sample sites.

3.11.1 Climatic variables

Climatic variables used in the JVMP included mean annual rainfall and mean annual temperature. These variables were the same as those used in the gap analysis described in section 3.1.3.

3.11.2 Edaphic variables

Edaphic variables used in the JVMP vegetation models included soil fertility and soil rooting depth. Edaphic variables related to soil were derived from the DLWC BBS bioregion Soil Landscape Reconnaissance Mapping Project (DLWC, 2002).

3.11.3 Physiographic variables

Physiographic variables included a digital elevation model (DEM) used to derive climatic surfaces for rainfall, temperature and wetness. The DEM, used for both the gap analysis and the modelling, is described in section 3.1.3.

3.12 MODELLING COMMUNITY DISTRIBUTIONS

The JVMP utilised Generalised Dissimilarity Modelling (GDM) carried out by the NPWS GIS Research and Development Unit (GIS unit) in Armidale.

3.12.1 General modelling strategy

The general strategy used to model vegetation group distributions in the JVMP is depicted in Figure 11. The strategy is based on a standard 'classification-then-modelling' approach. Ferrier, et al., 2002 discusses this approach in relation to other possible approaches.

Floristic survey sites were initially grouped into vegetation groups using numerical classification (as described in Section 3.9). The distributions of these groups were then modelled and extrapolated in relation to a set of mapped environmental variables (for examples of previous applications of this general approach see Moore, et al., 1991; Keith and Bedward, 1999; Ferrier, et al., 1999b). While many different mathematical techniques could have been used to model vegetation group distributions (for example, decision-tree modelling or neural networks) the JVMP used a technique based on generalised linear/additive modelling (GLM/GAM). The benefit of this type of modelling was that probability surfaces could be produced for individual vegetation groups, thereby facilitating subsequent application of API constraints in the manner described below.

The traditional approach to using GLM/GAM to model preclassified vegetation groups is to fit a separate GLM or GAM to the data for each group (ie, sites at which the group is recorded as either present or absent). While this approach has many strengths (Ferrier, et al., 2002), a potential weakness is that the models are fitted independently of one another – ie each model is based on the data for a single vegetation group, and ignores the data for all other groups. Each model is therefore fitted using a relatively small proportion of the total information contained in the data set. This can limit the power of such models, especially for groups occurring at a small number of sites (ie with small sample sizes).

The JVMP modelled all of the vegetation groups simultaneously by fitting a single multivariate model to the entire data set. The JVMP used a combination of generalised dissimilarity modelling (GDM, a new technique derived from GLM/GDM) and k-nearest neighbour modelling (described in Sections 3.12.2 and 3.12.3).

The initial combined application of GDM and k-nearest neighbour modelling to the JVMP groups in relation to abiotic environmental variables produced a set of probability surfaces, one for each vegetation group. This indicated that the probability of that vegetation group occurring in each one hectare grid cell within the bioregion was based purely on mathematically modelled relationships with the abiotic environmental variables. The accuracy of these predictions was assessed by cross validation of the survey data (described in Section 3.12.4). The probability surfaces were then adjusted (or 'conditioned') using all available vegetation mapping and air photo interpretation, in conjunction with expert-derived rules. The rules specified the JVMP communities that could potentially occur in each mapped vegetation or API class (described in Section 3.12.4). These 'constrained probability surfaces' were used directly to estimate areas of vegetation groups occurring in each of the planning units employed when considering land use options as part of the WRA. The constrained probability surfaces were also used to derive two different versions of a composite vegetation map for the region, in which each grid cell in the region was assigned to a single vegetation group (described in Section 3.12.5).

3.12. 2 Derivation of generalised dissimilarity model (GDM)

GDM is a recently developed statistical technique for modelling the biological dissimilarity (turnover in species composition) between pairs of survey sites as a function of the environmental and geographical separation of these sites (Ferrier, et al., 1999b; Faith and Ferrier, 2002; Ferrier, 2002; Ferrier, et al., 2002).

The basic analytical strategy of GDM is derived from that of permutational matrix regression (for example, Legendre, et al., 1994) which uses multiple linear regression to predict the dissimilarities in a site-by-site matrix (the response) as a function of distances in one or more independent (explanatory) matrices. In the application of interest here the response matrix contains biological dissimilarities between all pairs of survey sites calculated using the Bray-Curtis measure (Bray and Curtis, 1957). A site-by-site matrix is also prepared for each of the explanatory variables. For example, if one of these variables is mean annual rainfall, then a matrix is prepared in which each value is the difference in rainfall between a given pair of sites. Significance testing in matrix regression is performed by Monte Carlo permutation to overcome the problem of dependency between pairs of sites. For previous examples of the application of matrix regression to ecological data see Poulin and Morand (1999), Ferrier, et al. (1999a) and Duivenvoorden, et al. (2002).

GDM extends the technique of matrix regression to address two types of nonlinearity commonly encountered in ecological data sets:

1. nonlinearity in the relationship between ecological distance and observed biological dissimilarity is accommodated by fitting models using generalised linear modelling (McCullagh and Nelder, 1989) instead of ordinary linear regression;

2. variation in the rate of biological turnover along different parts of an environmental gradient is accommodated through automated nonlinear transformation of environmental variables, using I-splines (Winsberg and De Soete, 1997).

The first step towards modelling the distribution of JVMP vegetation groups was to apply GDM to the same site-by-species data matrix used in the PATN analysis described in Section 3.9. The compositional dissimilarity between pairs of survey sites (Bray-Curtis measure based on presence/absence of species) was modelled in relation to the following:

- mean annual temperature
- mean annual rainfall
- Prescott moisture index (Index of soil water balance, from Prescott, 1948)
- Rei250 - Relative Elevation Index (250m radius)
- soil fertility
- soil rooting depth
- distance to nearest water body or river and
- geographical separation.

Other environmental variables, also considered as candidate predictors, were subsequently excluded because they did not add significantly to the fit of the model once all of the above variables were included. These were:

- Rei250 - Relative Elevation Index (500 metre radius)
- Relative Elevation Index (1000 metre radius)
- soil water holding capacity
- soil drainage
- topographic position
- terrain-corrected solar radiation.

3.12.3 Extrapolating vegetation group distributions using k-nearest neighbour modelling

The transformed environmental space derived from the GDM provided the basis for extrapolating distributions of the JVMP groups generated by the PATN analysis (see Section 3.9). This extrapolation was performed using variable-kernel similarity metric (VSM) learning (Lowe, 1995), a form of k-nearest neighbour modelling.

For each community, *i*, the probability, p_i, of that group occurring at a given grid cell was predicted as a function of the observed occurrence of the community at the *J* survey sites nearest to (ie most similar to) the cell within the transformed environmental space:

$$p_i = \frac{\sum_{j=1}^{J} n_j s_{ij}}{\sum_{j=1}^{J} n_j}$$

where s_{ij} is the occurrence of vegetation group *i* at survey site *j* (0=absent, 1=present), and n_j is a distance weighting for site *j*, calculated as:

$$n_j = \exp\left[\frac{-d_j^2}{2r\left(\sum_{m=1}^{J} d_m \Big/ J\right)^2}\right]$$

where d_j is the distance (in transformed environmental space) between the grid cell of interest and survey site *j*, and *r* is a constant determining how quickly the weighting of sites declines with increasing distance. The values assigned to *J* (defined above) and *r* were optimised through cross validation, of the JVMP survey data (see Lowe, 1995 for details). Based on this cross validation, *J* was assigned a value of 80 and *r* a value of 0.5.

3.12.4 Constraining predictions using existing vegetation mapping and API

The predictions from the above modelling were further refined through integration with all available vegetation mapping and API within the region. An expert-derived technique described by Ferrier, et al. (2002) was used to constrain (or condition) predictions derived from modelling of these groups in relation to abiotic environmental variables.

The API composite layer (see Section 3.5) was used to constrain predicted vegetation group distributions. The process used a ‘look-up table’ specifying which JVMP groups the experts (ie the JVMP TWG) believed could potentially occur within each mapped class in the API composite layer.

The table was a simple cross-tabulation of API classes (rows) by vegetation groups (columns). Each cell of the table contained either a one, if the experts believed a given group could occur within a given API class, or a zero if they believed the group could not occur within the class.

Once prepared, the look-up table was used to automatically constrain the modelled probability surfaces for the JVMP vegetation groups. For each one hectare grid cell, any group with a zero entry in the look-up table for the API class mapped for that cell had its probability set to zero. The allowable groups (those with nonzero entries in the table) had their probabilities scaled upwards so that they summed to a total probability of one. This approach not only facilitated ready integration of modelling and API, but also allowed the relative weight given to the two data sources to be varied between different parts of the region. For example, if the experts felt that there was good correspondence between a given API class and a particular JVMP group then they could force the API to totally override the predictive modelling by placing a single one in the row for that class. At the other end of the spectrum, areas without any existing API were placed into a single ‘unmapped’ class with all ones in the relevant row of the look-up table. The predicted probabilities based on modelling remained unmodified in these areas.

The main drawback to this method was that only those vegetation groups which the model predicted could occur within a grid cell were able to be chosen by the experts, ie if the experts felt that a given API vegetation group would most likely be associated with a given vegetation group, derived from PATN analysis, the method would only allow that PATN group to have priority if the model had previously predicted it would occur within that grid cell.

3.12.5 Deriving a composite vegetation map

While the use of probability surfaces ensured that maximum information was available for each vegetation group, there was also a requirement to generate a composite (single layer) vegetation map for the purposes of communication and interpretation. Two different versions of a composite map were produced, each with particular advantages and disadvantages.

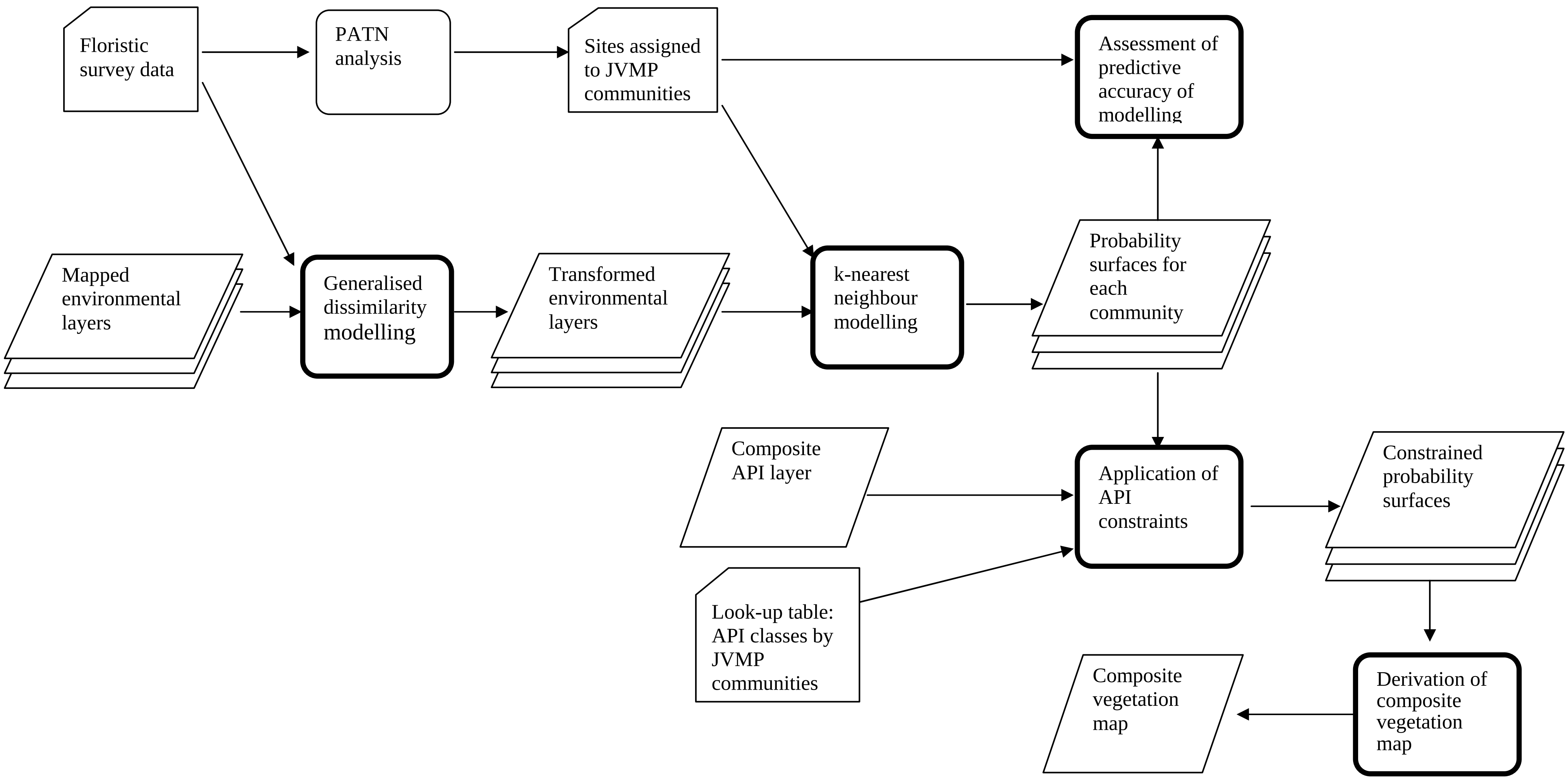
In the first version, each grid cell in the region was simply assigned to the JVMP vegetation group that has the highest predicted probability, based on the constrained probability surfaces.

While this composite provided the best indication of the group most likely to occur at any given location, it was possible to misrepresent the total extent of groups. In particular the extent of more common (or well sampled) groups was likely to be over estimated while the extent of rarer (or poorly sampled) groups was likely to be under estimated (some less well sampled groups were not represented at all in this composite). For example, if in a given part of the region, Vegetation group A was predicted to occur with a probability of 0.6 and Vegetation group B with a probability of 0.4 then the composite would depict Vegetation group A as occurring across this whole area. The information on other less-probable groups would have been lost.

The second version of the composite was derived using an iterative Bayesian technique (Strahler, 1980; Ferrier, et al., 2002) in which the total extent mapped for each group was matched as closely as possible to the total extent predicted from the probability surface for that group (ie by summing the grid cell probabilities). This composite provided users with an indicative representation of the extent of each group across the region. However, the vegetation group mapped for each grid cell was no longer necessarily the most likely type occurring at that cell. The probabilities of less well sampled groups were artificially inflated to allow those types to ‘appear’ through the more common groups.

-

Figure 11. General strategy used to model vegetation group distributions



3.13 PRODUCT INTEGRATION

3.13.1 Potential vegetation distribution models of the BBS

One hundred and fifteen probability surfaces were produced, each one representing an individual vegetation group as derived through PATN analysis. Vegetation groups were represented in the composite map according to their areal extent derived from the individual probability surfaces. This method provided greater utility than a simple allocation based on highest probability. This was because, in instances where the ratio between the predicted gross and net area is large, multiple vegetation groups may be predicted for each grid cell and are more likely to be captured in the composite model using the Bayesian technique (Section 3.12.5).

3.13.2 Current extant vegetation models of the BBS

The DLWC landcover layer and the API composite layer were used to create a mask. An extant vegetation model (in map form) was produced by applying the mask to the composite map and each individual probability surface. Table 7 presents a summary of the information contained in the DLWC landcover data set (originating from satellite image interpretation of Landsat 7 TM datasets).

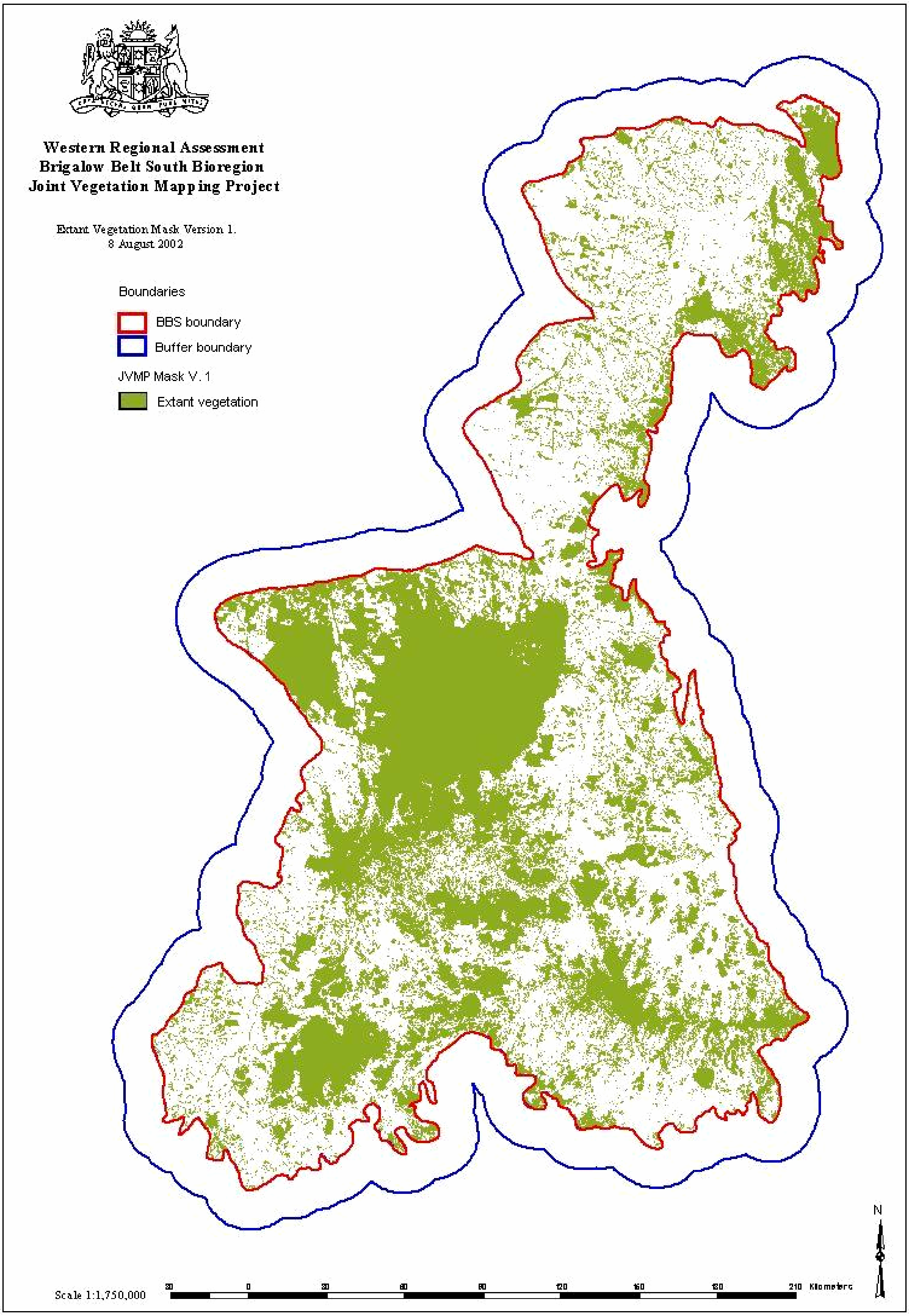
TABLE 7: SUMMARY INFORMATION FROM DLWC BBS BIOREGION LANDCOVER DATA

Landcover class	Area (ha)
Timber (greater 15% ccp)	1 556 843
Water	14 350
Cropping	1 540 575
Wetlands	1 222
Urban	6 543
Open woodland / grassland	2 130 901

Initially the composite API layer (described in section 3.5) was supplemented by the addition of the vegetation layer identified as ‘timber’ within the BBS bioregion landcover dataset (derived by DLWC).

The landcover dataset identified areas within the landscape according to key vegetation and landuse features. For example, areas nominated as ‘timber’ were identified, through satellite image interpretation, as having greater than or equal to fifteen percent crown canopy cover. The timber layer was merged with the composite API layer and was used to fill in data gaps in the composite layer. This ensured that there were as few gaps in the ‘woody’ component of the extant mask as the available data allowed. Figure 12 illustrates the landcover of the original extant vegetation mask, which resulted in the identification of a probable 2 214 995 hectares of woody vegetation.

Figure 12: Extant vegetation mask version 1



After comparing the mask with the distribution of the survey sample sites it was discovered that 547 sites, out of the 3078 survey sites utilised in the analysis, occurred within the open woodland/grassland layer. Of these sites, 311 occurred outside the masked area. This equated to about 10% of all survey sites being excluded from areas defined as having extant native vegetation.

The 547 survey sites represented 76 vegetation groups or 66% of the identified vegetation groups. A threshold was set where if more than 20% of a vegetation group’s total number of plots were recorded as falling within the identified open woodland/grassland, then that vegetation group would be considered as having an open woodland/grassland component. The project identified that 34 (or 29.5%) of the identified vegetation groups had an open woodland/grassland component. The 547 sites represented 61.8 % of the 731 sites that contributed to the derivation of the 34 open woodland/grassland vegetation groups.

As a result of this analysis a decision was made to incorporate the vegetation represented by the 34 vegetation groups into the extant mask. The following steps outline the process used to extend the extant mask into the open woodland/grassland areas, as identified by the DLWC landcover data layer.

Step 1. The open woodland/grassland areas were selected from the DLWC landcover dataset, 2 130 901 hectares.

Step 2. The extant mask v.1. was then subtracted from the open woodland/grassland data layer. This resulted in the identification of the nonassigned open woodland/grassland vegetation, 1 593 680 hectares.

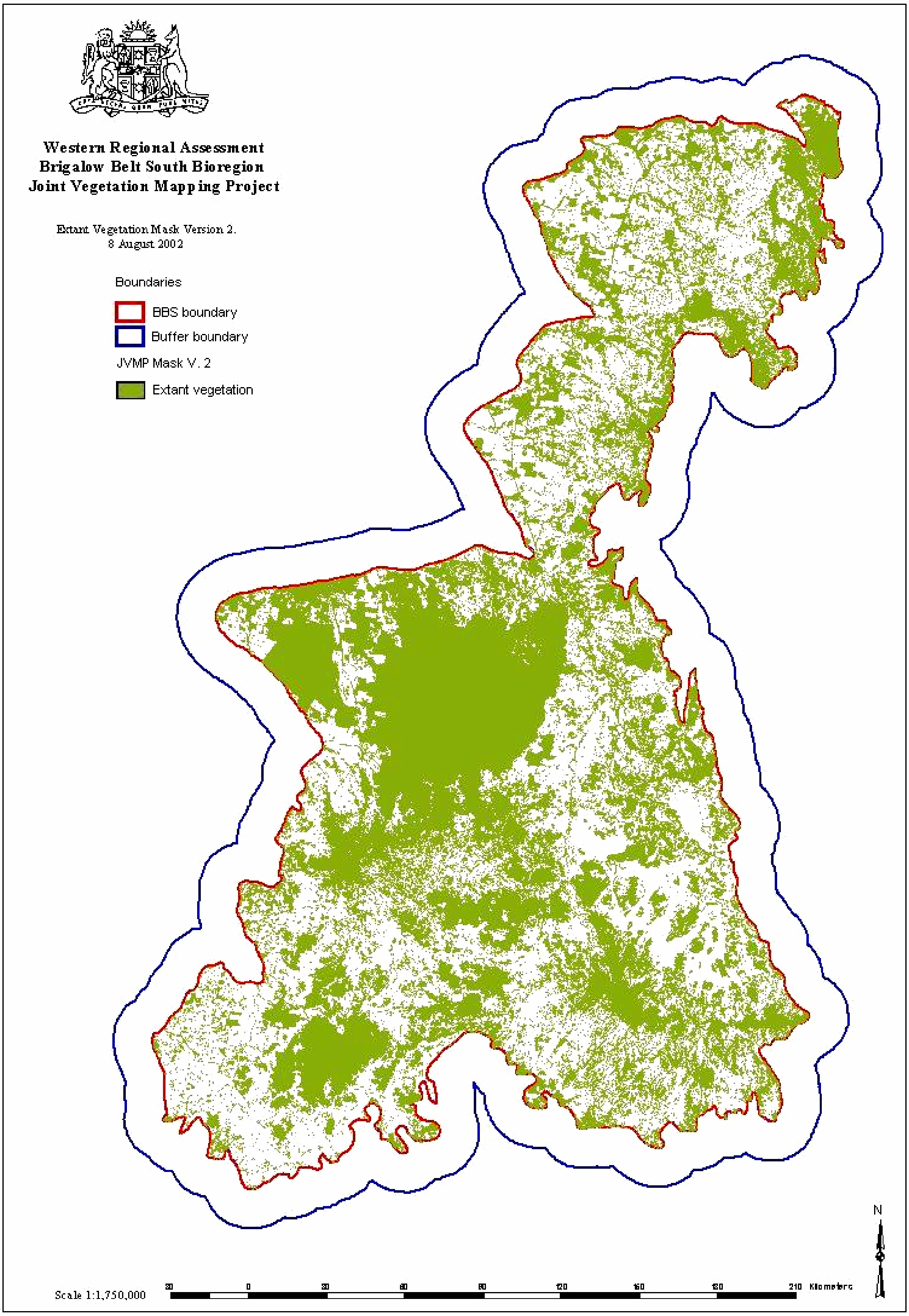
Step 3. The 34 vegetation groups were selected from the composite potential vegetation distribution layer. This was done by reclassing the vegetation groups, which were not represented by the 541 aforementioned survey sites, with the resultant distribution of the 34 target vegetation groups being accepted as their likely extent, 1 664 018 hectares.

Step 4. The layers selected in steps 2 and 3 were intersected to define the extent of the additional native vegetation, as defined by the 541 survey sites, which occurred in the non assigned open woodland/grassland landcover data, 524 868 hectares.

Step 5. The area defined in step 4 was added to the original mask to derive the final extant vegetation mask, 2 739 814 hectares (Figure 13).

This method overcame the limitations of using only the derived DLWC open woodland/grassland data layer because it incorporated the restraint of the modelled distribution. Instead of an additional 2 130 910 hectares being incorporated into the extant vegetation mask less than 25% of that figure was incorporated. This additional 524 868 ha was considered to be more reflective of the true extent of the extant open woodland / grassland component of the native vegetation of the BBS bioregion.

Figure 13: Extant vegetation mask version 2



4. RESULTS

4.1 GAP ANALYSIS

The JVMP selected a new set of candidate sites for each of the gap analyses conducted. Ten thousand candidate sites were randomly selected for Gap Analyses 1 to 3 and 20,000 sites were randomly selected for Gap Analyses 4 to 6. This ensured that each iteration of the gap analysis was treated as a separate survey and that each candidate site was afforded the highest priority of selection at each iteration. (Refer to Section 3.1 and Appendix 11).

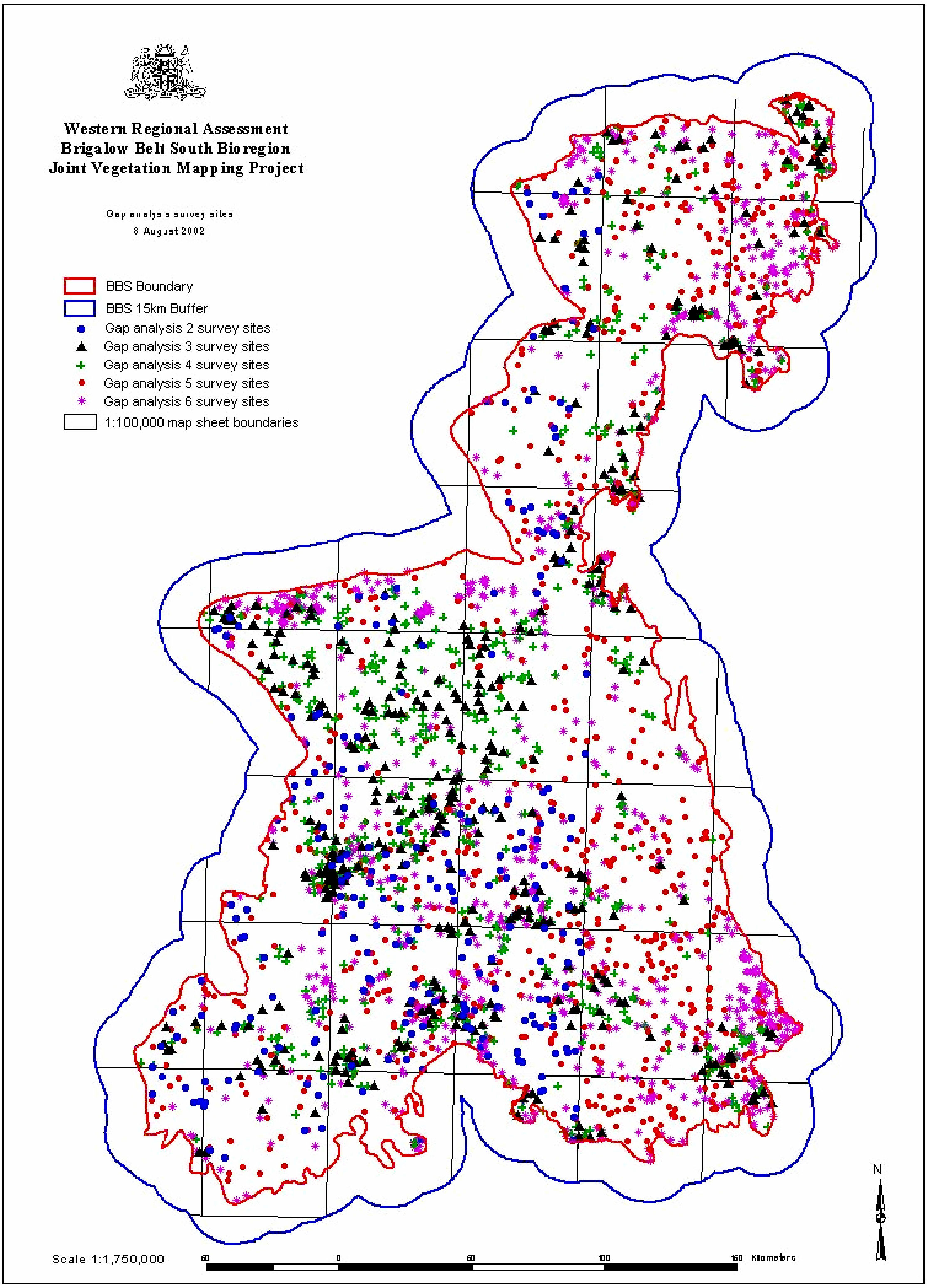
The outputs from each of the gap analyses are provided in Figure 14. The spatial distribution of the sites across the bioregion in relationship to the candidate and existing survey sites is evident. The distribution of the priority survey sites was uniform across the bioregion as illustrated in Figure 14. Each gap analysis provided a priority listing for each selected survey site (Method, Section 3.1.1), which enabled the botanists to plan their field work and guide survey site selection when access constraints applied.

Due to the unknown nature of the vegetation cover within the bioregion and the high cost of survey the TWG selected more survey sites than necessary at each iteration of the gap analysis. This provided a pool of survey sites, listed in order of priority, in case native vegetation was not present within the selected site locality or if landholders declined permission to access their properties. The number of survey sites generated by each gap analysis and date of selection was:

Gap 1	48 sites	July 2001
Gap 2	120 sites	August 2001
Gap 3	400 sites	September 2001
Gap 4	600 sites	October 2001
Gap 5	1000 sites	November 2001
Gap 6	1000 sites	December 2001

The gap analysis survey sites which had a woody vegetation mask applied were likely to be biased to the more extensively wooded parts of the landscape because the woody mask was applied after candidate sites were selected. Thus, the most extensively cleared groups were originally allocated a lower sampling intensity resulting in an inadequate definition of the vegetation groups present at those locations. Subsequently those vegetation groups of limited distribution were masked in the floristic analysis.

Figure 14. Gap analysis priority survey sites



4.2 FLORISTIC SURVEY

4.2.1 JVMP survey

Two hundred and forty floristic sites were sampled within the Boggabri and Blackville 1:100 000 map sheets. The NVMP survey yielded 103 sites in the Bellata and 186 sites in the Gravesend map sheets. In other words, 529 sites were surveyed before delivery of gap analysis outputs.

Seven hundred and twelve sites were completed after the gap analysis sites were delivered. In all, the JVMP floristic survey produced 1241 new survey sites across the BBS.

Examination of the survey site locations at the completion of the survey stage revealed a bias in the sampling. Survey sites in the south of the BBS bioregion exhibited a bias towards crown lands. This was due to a number of factors including access constraints to privately managed property; resulting in a focus on the crown estate by some botanists.

Figure 15 illustrates the regression analysis carried out to determine the relationship between the number of survey plots and the area of each vegetation group as predicted by the model (further explained in section *4.5.2 Predicted potential vegetation distribution*). The associated statistics provide information on the fit of the regression line to the data. The multiple R^2 value of 0.705 indicates that there is a strong relationship between the number of plots in each vegetation group and the net area predicted by the model.

Figure 16 illustrates the distribution of the survey sites selected for sampling as a result of gap analysis. Figure 17 illustrates the location of completed survey sites. From comparison, it was evident that some priority areas were not sampled. The level of bias evident in the sampling was not determined, although of the 3168 sites selected for survey by the gap analysis tool 712, or 30% were completed.

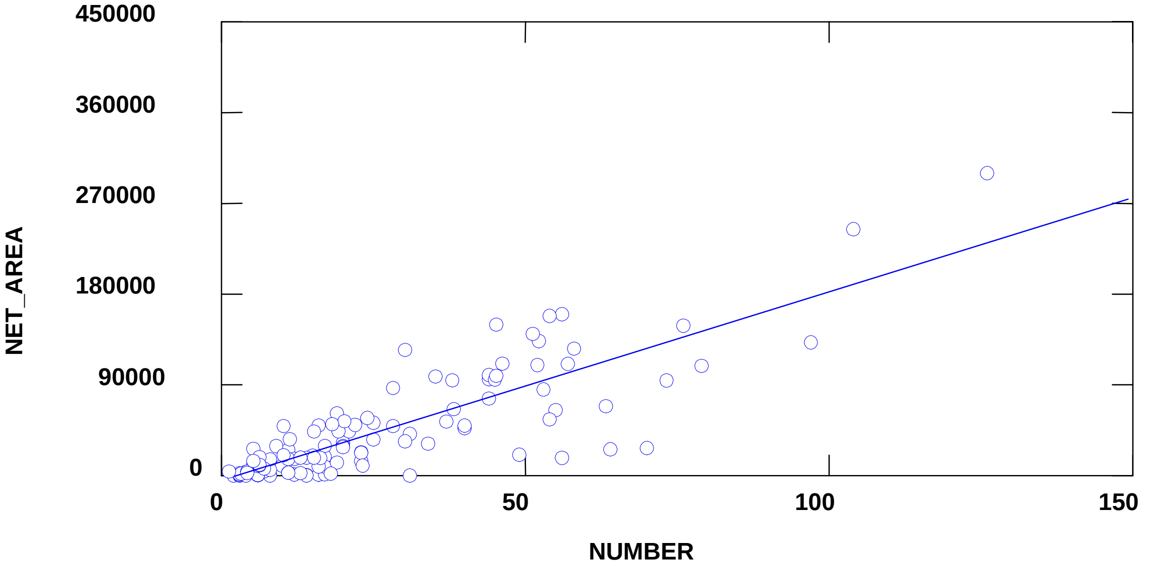
It is likely that some of the modelled groups may have been over modelled. This may have affected the results of modelled vegetation groups.

If a group is over modelled, the model will predict a distribution greater in area than would be likely to occur within the landscape. This could markedly reduce the usefulness of the potential distribution model. This issue was raised with the WRA Steering Committee and additional funding was approved to carry out floristic survey on privately managed lands. Unfortunately, due to time constraints, the additional survey work was not conducted and the bias towards public lands in the floristic sample remained.

4.2.2 Combined floristic data set

The combined floristic data set resulted in 3 166 survey sites available for analysis and this resulted in one survey site for every 1 658 hectares within the BBS.

Figure 15: Regression analysis for number of plots and area for each vegetation group as predicted by the model — R^2 value of 0.705



6 case(s) deleted due to missing data.
Dep Var: NET_AREA N: 115 Multiple R: 0.840 Squared multiple R: 0.705

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
Constant	-5378.599	4046.690	0.000	.	-1.329	0.186
Number	1866.646	113.505	0.840	1.000	16.446	0.000

Figure 16: Gap analysis survey sites from gap analysis 2 to 6

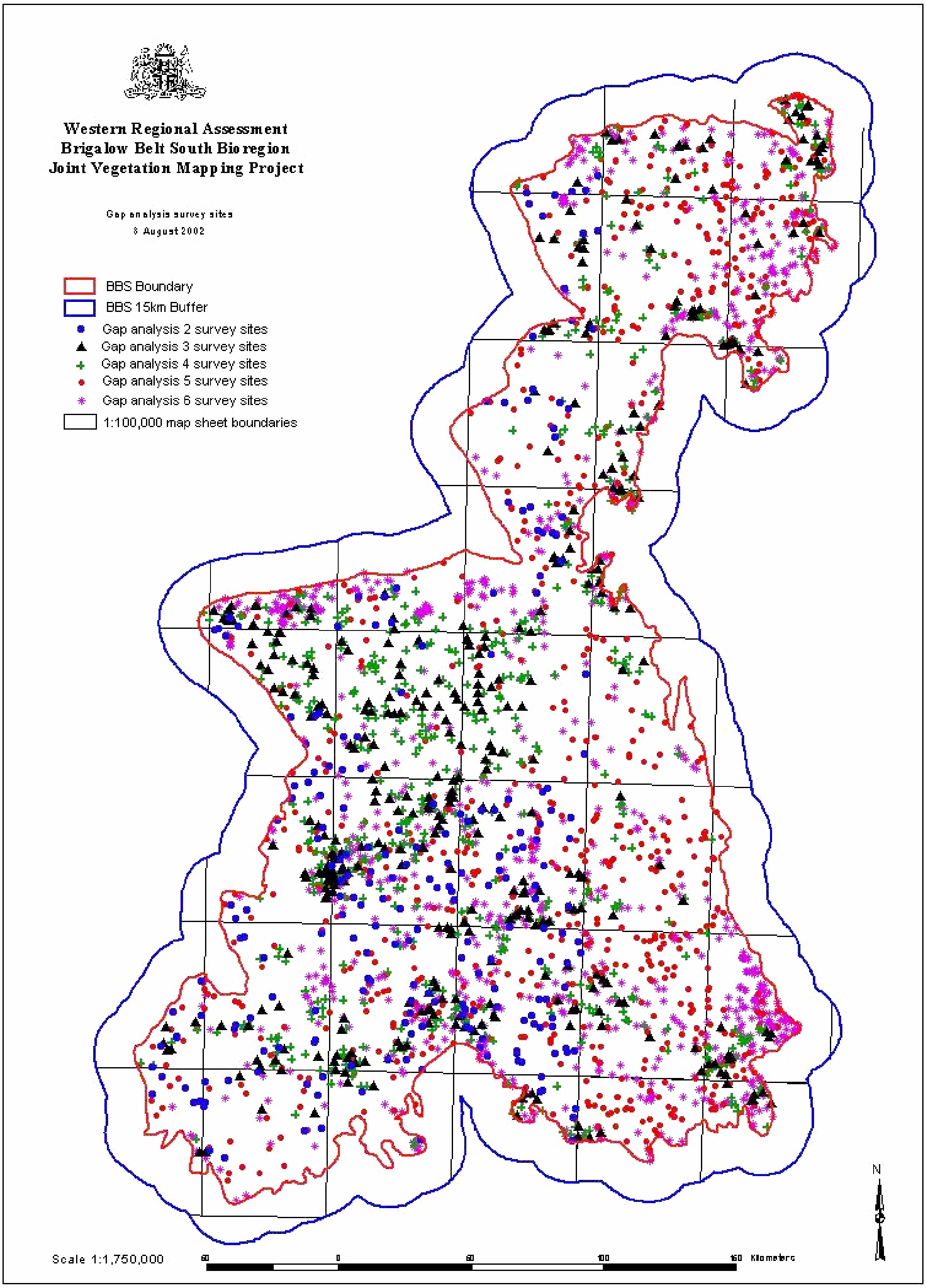
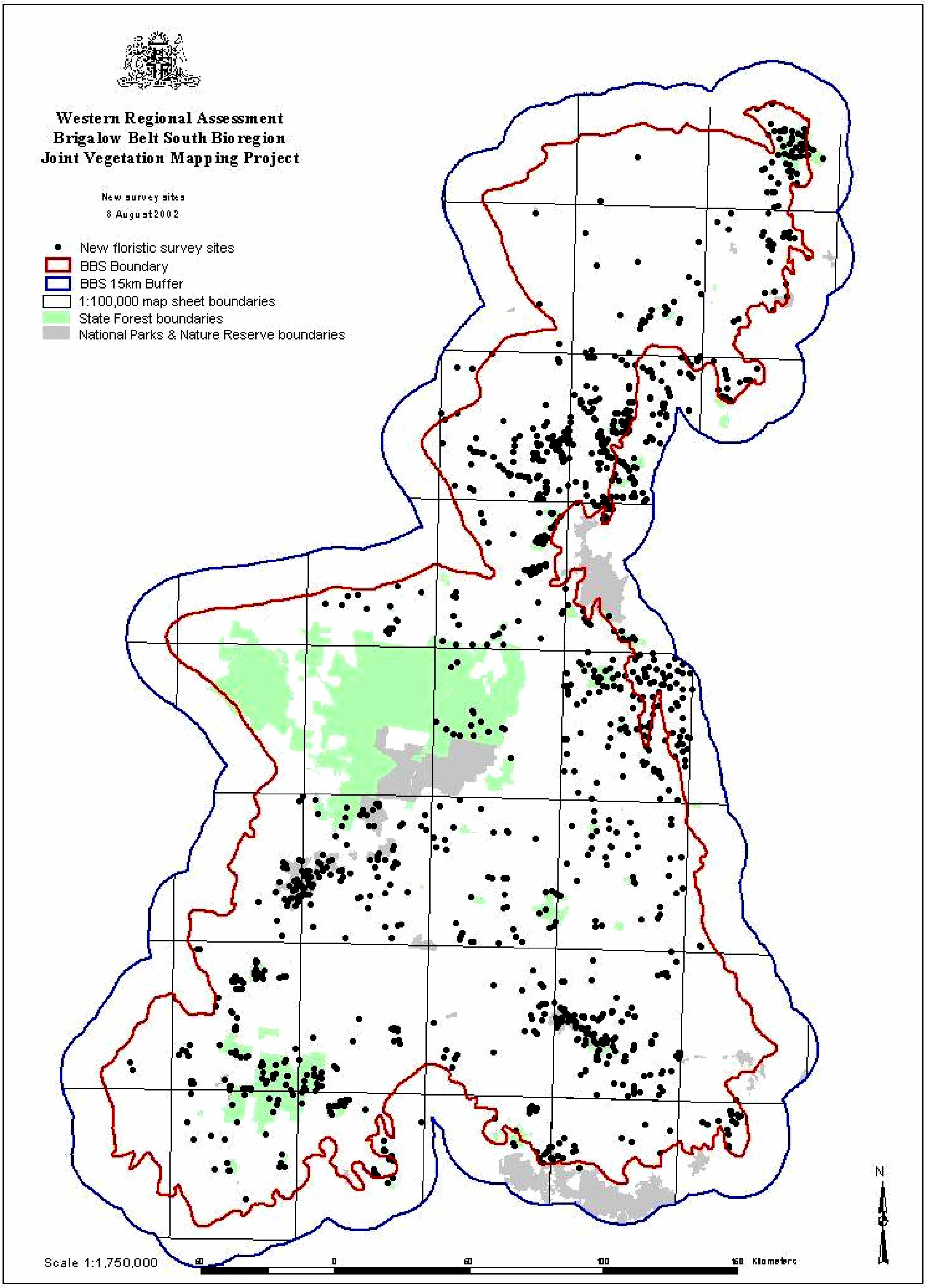


Figure 17: New survey sites as a result of JVMP survey effort



4.3 AERIAL PHOTOGRAPHY INTERPRETATION

4.3.1 Targeted API

The targeted API provided information on the structural characteristics and overstorey composition of polygon elements across the landscape. Available aerial photography was restricted to 1:50 000 and in limited instances 1:25 000 scale. Analysis of the structural information derived from the API indicated that the structural information was not consistent for all targeted map sheets. As a result only the overstorey composition of the polygon elements was utilised in the derivation of the composite API data layer.

The targeted API program was unable to deliver all of the required map sheets determined by the TWG. As a result there is little current spatial information within the Merriwa map sheet. The Merriwa map sheet was poorly sampled and the modelled potential vegetation distribution map may be less reliable in this area.

Additional targeted API was carried out within the Bingara and Mendooran map sheets. The Bingara API was restricted to the State forests within the sheet. The Mendooran API was carried out in lieu of NVMP API, instead focussing on the woody vegetation component.

Some targeted API polygons were not attributed. These polygons had little utility when used as a constraint on the model. Figure 18 illustrates the extent of the targeted API carried out for the JVMP. Table 8 provides data on the area of woody vegetation identified within each map sheet and the extent of unattributed polygons by map sheet.

TABLE 8: AREA OF WOODY VEGETATION BY MAP SHEET FOR TARGETED API MAP SHEETS AND PERCENTAGE OF UNATTRIBUTED WOODY VEGETATION POLYGONS

Map Sheet	Woody veg (ha)	Unattributed Woody veg (ha)	Percentage unattributed
Bingara	5 478.9	162.273	3.0%
Blackville	10 1781.799	4 559.288	4.5%
Cobbora	32 744.593	3 284.158	10.0%
Gulgong	17 611.641	7 072.959	40.2%
Mendooran	72 151.645	111.908	0.2%
Murrurundi	29 535.749	835.46	2.8%
Yallaroi	54 182.503	0	0.0%
Yetman	70 788.163	19 338.092	27.3%

4.3.2 NVMP API

The JVMP delivered five of the six NVMP map sheets nominated in the project proposal. These were the Boggabri, Curlewis, Coonabarabran, Gravesend and Tambar Springs 1:100 000 map sheets. Each was completed to NVMP technical standards as detailed in the DLWC *Guidelines for mapping native vegetation*. Figure 19 illustrates the completed NVMP map sheets.

Figure 18: extent of targeted API for the JVMP

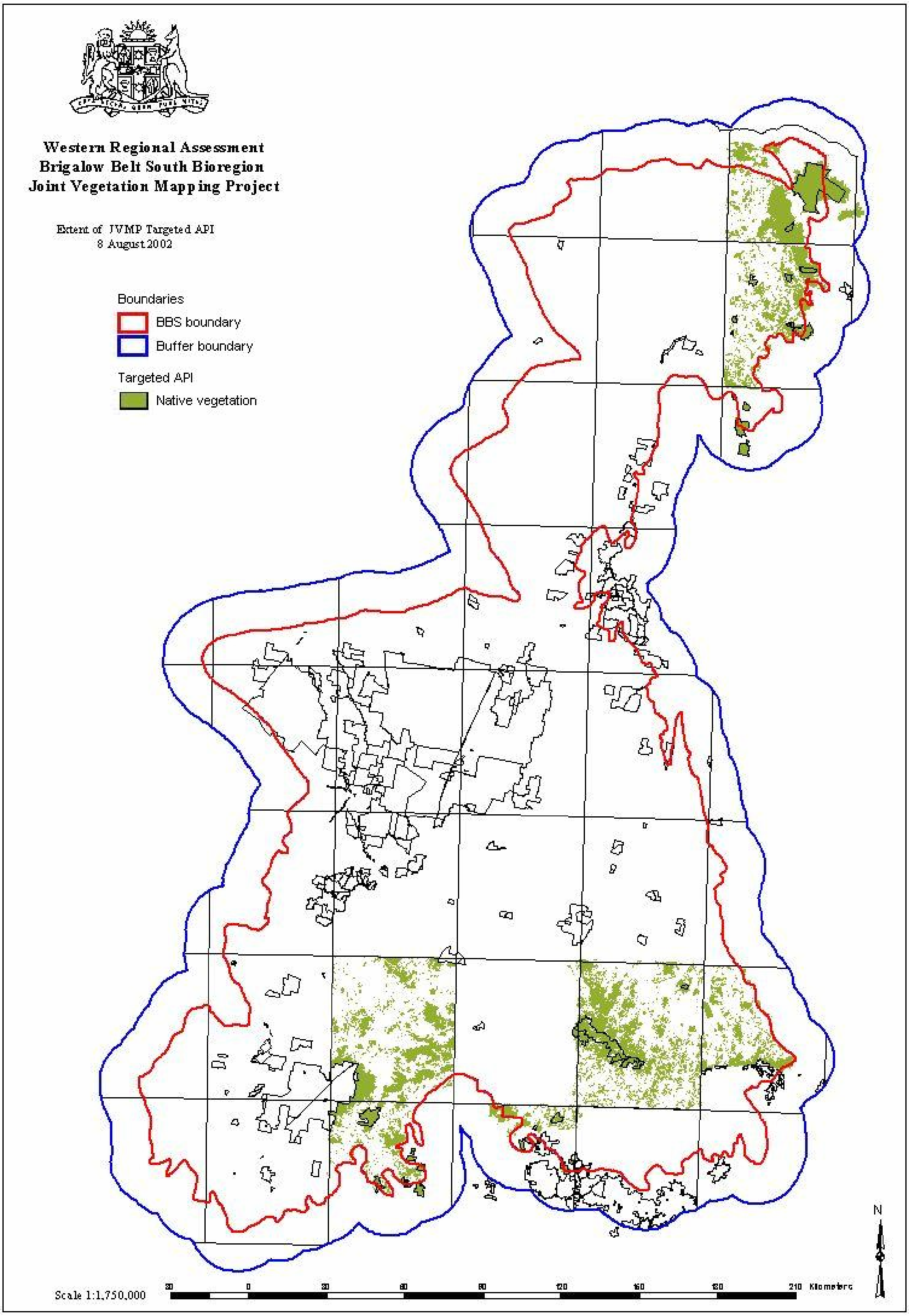
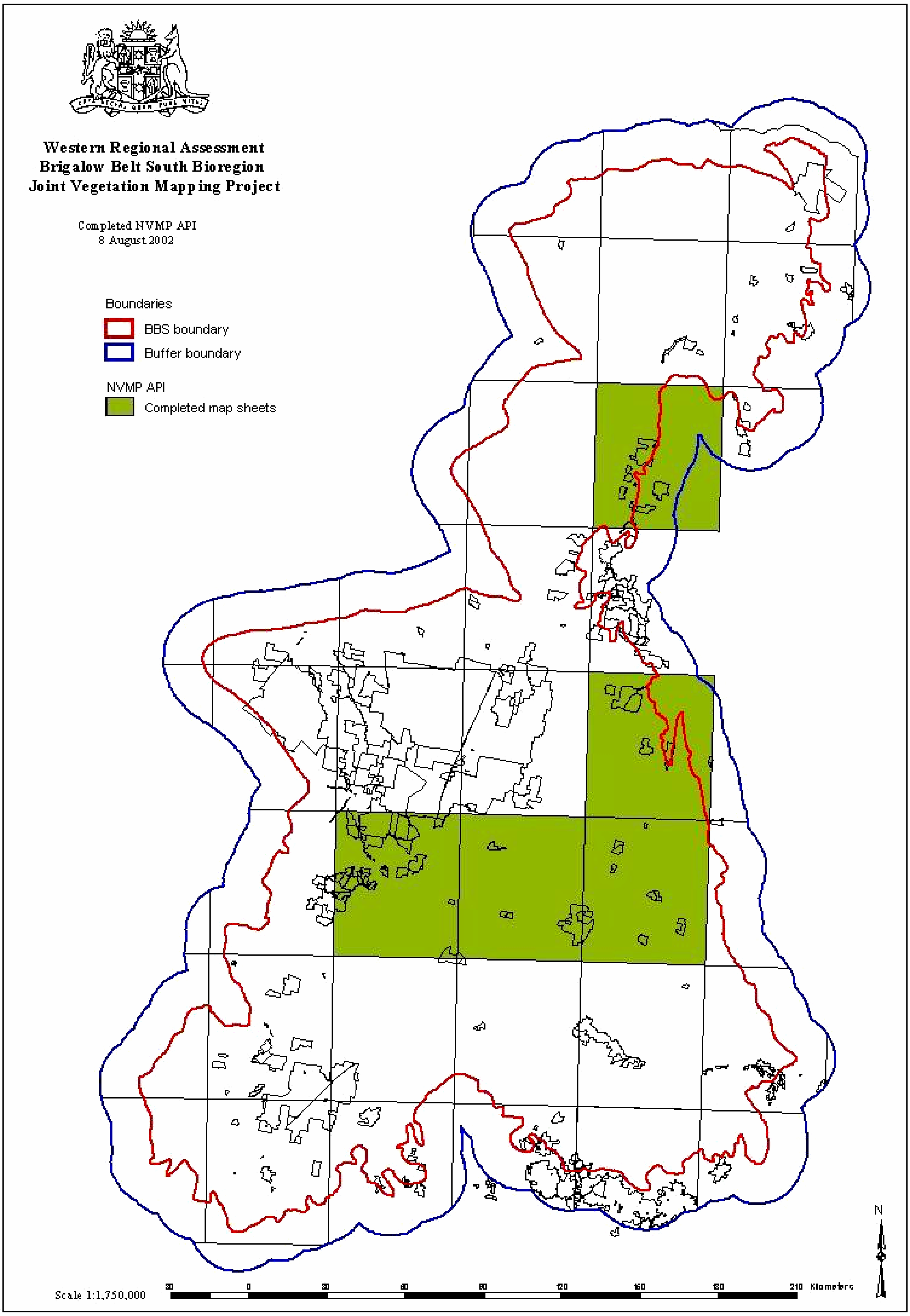


Figure 19: Completed NVMP map sheets within and neighbouring the BBS bioregion



4.3.3 Composite API layer

All woody vegetation API data sets available to the JVMP were amalgamated into a composite API layer. In all, the API composite layer used 31 API data sets. Of these, 19 data sets were less than two years old; primarily NVMP API or JVMP targeted API. The composite API layer sampled a total of 2 013 709 hectares of woody vegetation. The extent of native vegetation derived by the composite API layer is illustrated in Figure 20.

4.4 DATA ANALYSIS

The PATN analysis of the standardised data set resulted in the identification of 115 native vegetation groups within the BBS bioregion and 15 kilometre buffer. Vegetation group descriptions are provided in Appendix 3 for all identified vegetation groups. Environmental, floristic and structural profiles have been prepared for each group. Diagnostic species and species with high frequency or high median cover values are documented in addition to all tree species with a cover score greater than 5%.

4.4.1 Floristic analysis and aggregation of vegetation groups

The final data matrix used for analysis comprised 3139 sites and 1155 taxa. The homogeneity analysis suggested that between approximately 50 and 100 groups were necessary to adequately characterise the floristic variation in the data. An initial level of 200 groups was selected from which to develop the final set.

The nearest neighbour check of the results showed that a high proportion of sites were misclassified, based on the chosen criteria. Table 9 summarises the results of the series of reallocations. Note that in this summary, it is possible for the number of reallocations to exceed the initial number of misclassified sites as some sites are reallocated more than once during iteration. in all cases, the process stabilises quickly after four iterations or less.

Forty-nine sites were not allocated to any group due to unresolved and ambiguous relationships. Most of these were highly disturbed sites or transitional sites. Further sampling might have confirmed whether any unallocated groups had represented distinct floristic assemblages. Following reallocations of misclassified sites, 19 of the initial groups disintegrated, having all of their constituent sites reallocated to other groups. These were all small groups of seven sites or less.

A further 66 groups were considered to be artefacts of disturbance, observer or season and were reallocated. The final 115 groups are listed with allocation history in Appendix 3.

The mean within-group dissimilarity (0.684) was significantly (t test, p=0.04) less than the alternatives of hierarchical classification to 115 groups (UPGMA using Bray-Curtis association, mean dissimilarity 0.696), and non-hierarchical classification to 117 groups using the PATN module ALOC with Bray Curtis association measure (mean dissimilarity 0.698).

However, the weighted mean dissimilarity (0.693) was higher than the weighted means of the alternatives (0.690 and 0.685 respectively). The difference was due to the influence of several relatively heterogeneous groups, formed by merging similar initial groups, especially the aggregate group 35 which was formed from merging several groups judged to be unduly influenced by disturbance.

Despite the aggregation of groups thought to be artefacts, it is likely that at least some of the remaining groups also represent, or are strongly influenced by, artefact. Some groups and some environments were very poorly sampled and are thus poorly characterised. Further sampling would reveal structure within some of these and would elucidate relationships among them.

The characteristics of the 115 groups are summarised in Appendix 3.

Figure 20: Extent of native vegetation as defined by the composite API layer

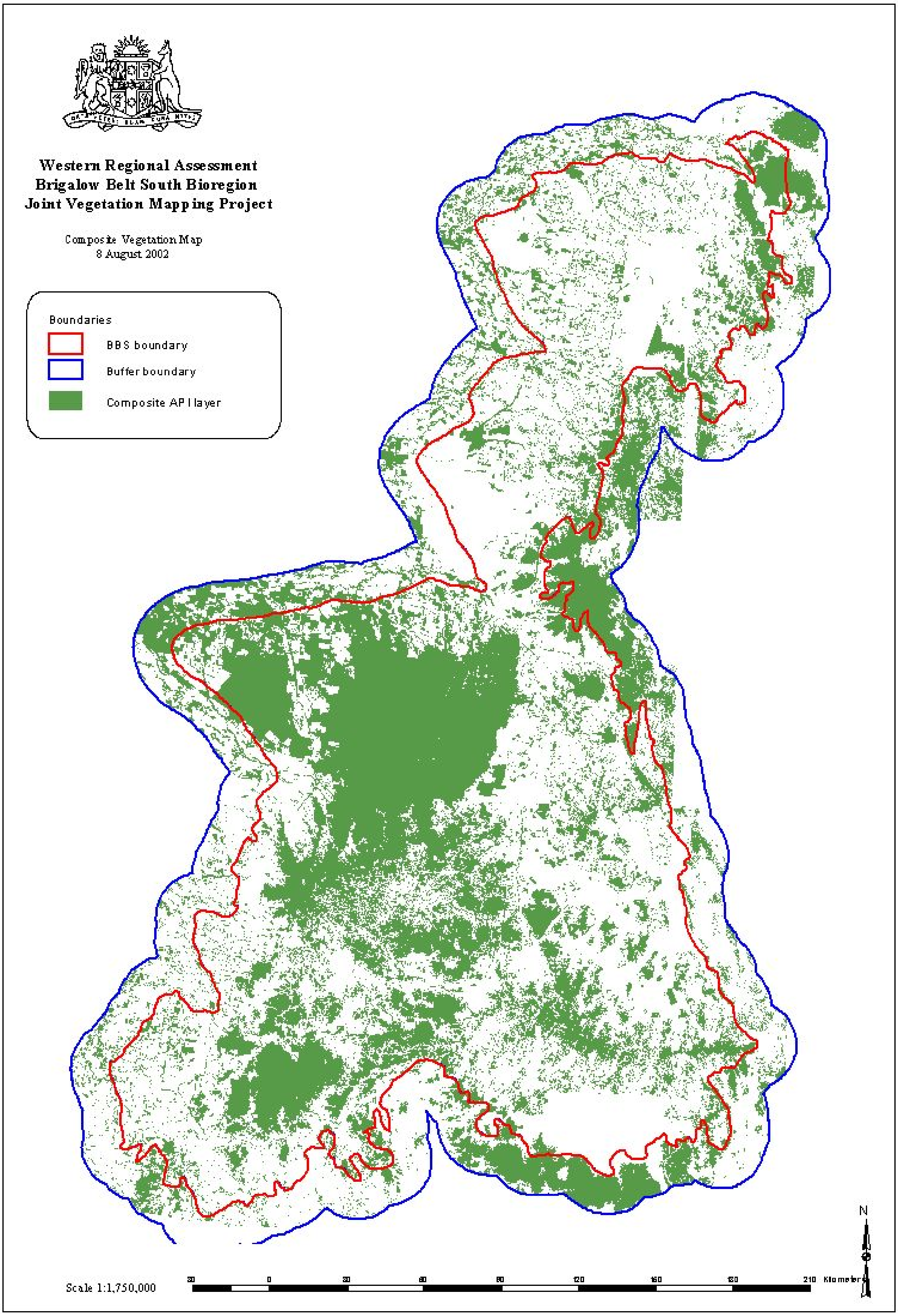


TABLE 9. SUMMARY OF RECLASSIFICATION RESULTS

Misclassification criterion	Initial number of misclassified sites	Number of reallocations	Number of iterations	Residual number of misclassified sites
All 5 nnbs in group other than site group	106	12	2	96
At least 4 nnbs in group other than site group AND most similar nnb in different group	287	52	2	249
At least 3 nnbs in group other than site group AND two most similar nnbs in group other than site group	248	91	3	174
At least 3 nnbs in group other than site group AND two most similar nnbs in group other than site group	174	35	2	142
At least 3 nnbs in group other than site group AND the most similar nnb in group other than site group	142	169	2	0

1. nnbs — the five nearest neighbours to each site.
2. Note that in this summary, it is possible for the number of reallocations to exceed the initial number of misclassified sites as some sites are reallocated more than once during iteration. In all cases, the process stabilises quickly after four iterations or less

4.5 MODELLING

Potential vegetation distribution models were developed for each identified vegetation community within the BBS bioregion. Model outputs were in the form of probability surfaces with a 100 metre grid cell resolution. Within each probability surface, each pixel has a unique value of between zero and one. Zero represents the lowest level of probability that can be reached and one represents the highest level of probability. Generally, probability values will be somewhere between zero and one except for instances where the composite API layer is constraining the model and then probabilities of one will occur.

The fitted GDM explained approximately 59% of the deviance (variation) in observed floristic dissimilarities between sites. Transformations (I-spline functions) fitted to each of the environmental variables are depicted in Figure 21. By transforming the environmental layers according to these functions, a transformed multivariate environmental / geographical space was generated to best fit the observed pattern of floristic dissimilarities within the region. The transformed layers were derived and stored at one hectare grid resolution. A scatter plot depicting the fit of predicted ecological distances (from the GDM) to observed dissimilarities is presented in Figure 22.

4.5.1 Evaluating predictive accuracy of modelled distributions

The predictive accuracy of the modelled distributions was evaluated using a form of cross-validation or jackknifing (Efron and Tibshirani, 1993; Pearce and Ferrier, 2000). This involved withholding each of the JVMP survey sites in turn from the modelling process, fitting a model to the remaining sites, then comparing the predictions obtained for each withheld site with the actual community recorded at that site.

The results of the evaluation of accuracy based on cross-validation are presented in Figures 23 to 26. These results describe the average performance of predictions across all vegetation groups combined. A detailed evaluation and comparison of the performance of modelling for individual groups was beyond the scope of this project. However, given the current availability of rigorous analytical techniques (see Stehman and Czaplewski, 1998; Pearce and Ferrier, 2000) for evaluating probabilistic predictions such as those generated by the JVMP modelling, this more detailed evaluation would be worth pursuing in the future (particularly if applied to independently collected survey data, as suggested above). In the meantime, it is hoped that the results presented in Figures 23 to 26 will instil users with at least some confidence in the utility of the modelled distributions.

Figure 23 depicts the relationship between the predicted probability of occurrence of a given vegetation group at a withheld site (based on the respective jackknifed model) and the observed proportion of sites at which the predicted group actually occurs (for more detail on this evaluation technique see Pearce and Ferrier, 2000). The predicted probabilities are grouped into 0.05 interval classes. Each of these classes includes data pooled from all groups. The graph suggests a reasonably close match between predicted probabilities of occurrence and observed proportions of sites occupied.

Figure 24 depicts the relationship between the number of survey sites at which a given vegetation group is predicted to occur (derived by summing the probabilities of occurrence predicted by the jackknifed model for each site) and the actual number of sites at which that group was recorded in the survey data set. Each symbol represents a different vegetation group. The close match between predicted and observed numbers of sites suggests that the probabilities predicted by the modelling provide a reasonable basis for estimating the total number of sites (or grid cells) at which a vegetation group is likely to occur within the region, or a defined part of the region.

Figures 25 and 26 show spatial variation in predictive accuracy across the region. Figure 25 depicts the ‘sum of squares’ (a measure of discrepancy between observed and predicted values) for each survey site, calculated by withholding this site from a jackknifed model based on the remaining sites. A higher sum of squares value indicates a higher discrepancy between observed and predicted occurrence. Figure 26 depicts a rough extrapolation of the sum of squares expected across all grid cells in the region. This is based on a simple linear regression of the sum of squares calculated for each survey site against the density of other survey sites around the site (in terms of the transformed environmental / geographical space employed in the modelling). In other words, site survey density was used to broadly indicate the accuracy of predictions derived from these sites.

Figure 21: Transformations (I-spline functions) fitted to each of the environmental variables included in the generalised dissimilarity model (GDM)

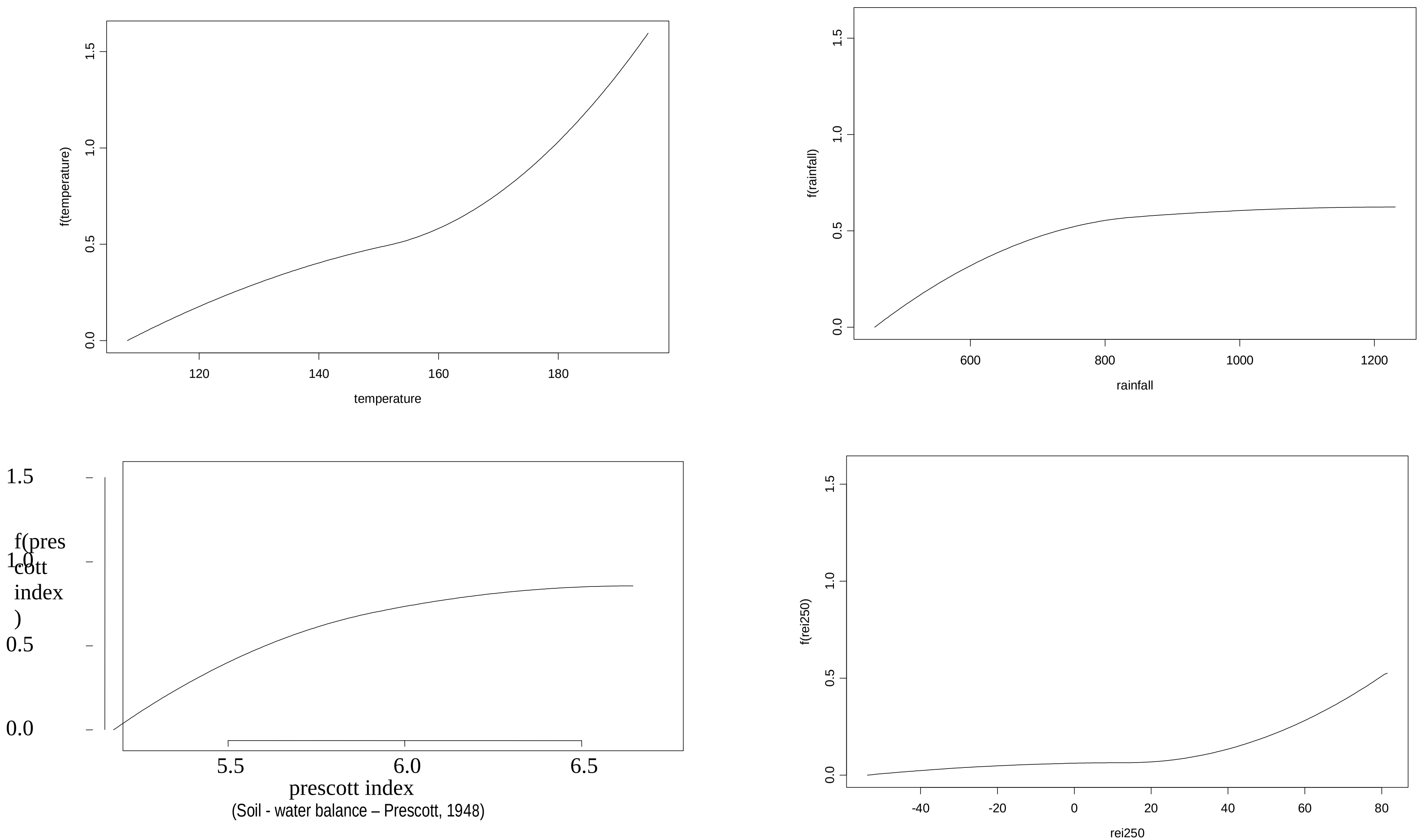


FIGURE 21.
(CONTINUED)

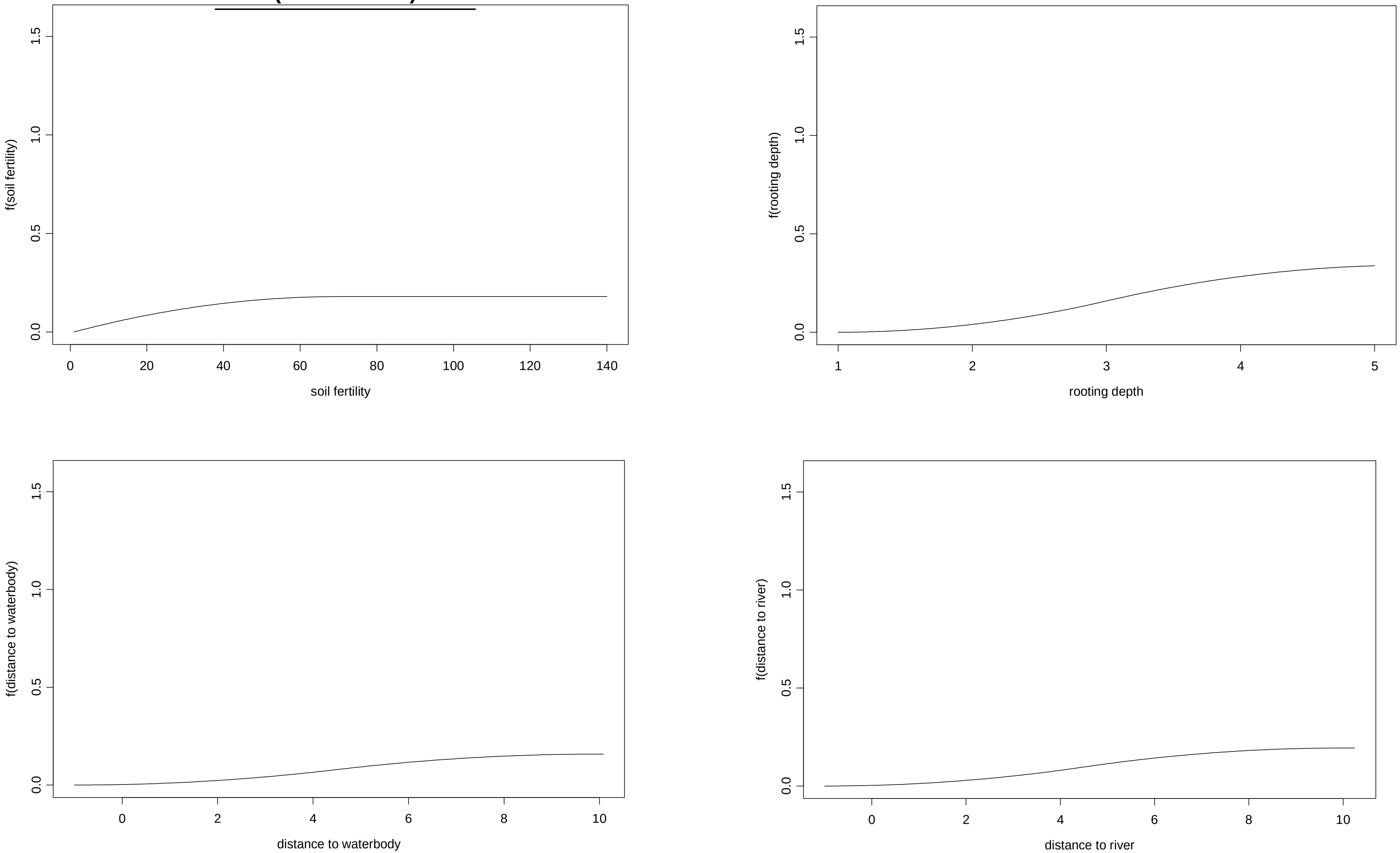


Figure 22: Relationship between predicted ecological distance, derived from the GDM, and observed floristic dissimilarity

Each dot represents a pair of survey sites. The curved line represents the link function employed in the GDM.

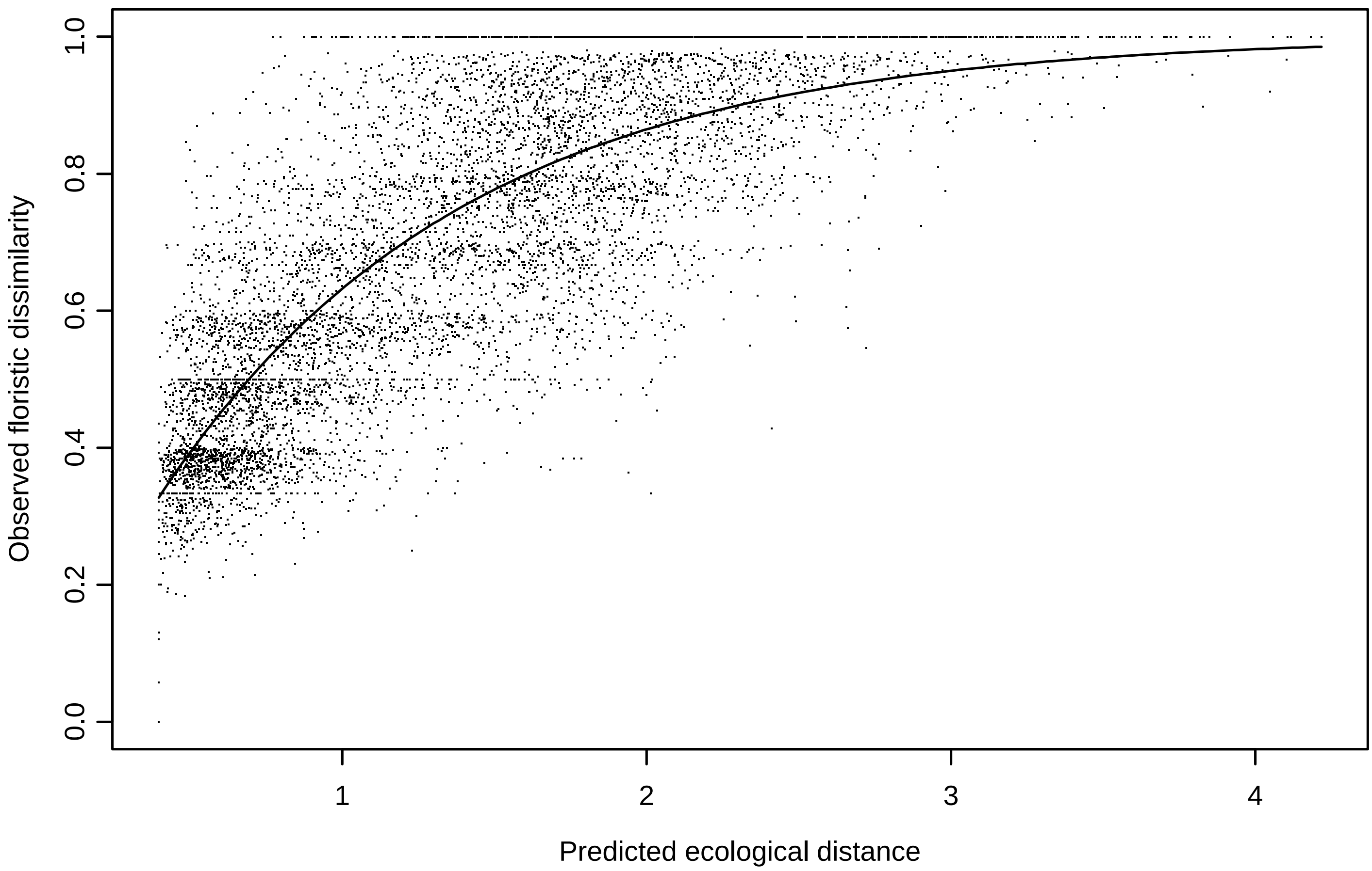


Figure 23: Relationship between predicted probability of occurrence of a given community (based on a jackknifed model) and observed proportion of sites at which that community actually occurs

The predicted probabilities are grouped into 0.05 interval classes. Each of these classes includes data pooled from all communities. The dot plotted for each class represents the observed proportion of occurrences, while the vertical bar represents the 95% confidence interval for this proportion. The diagonal line represents the relationship expected if predictions matched observations exactly.

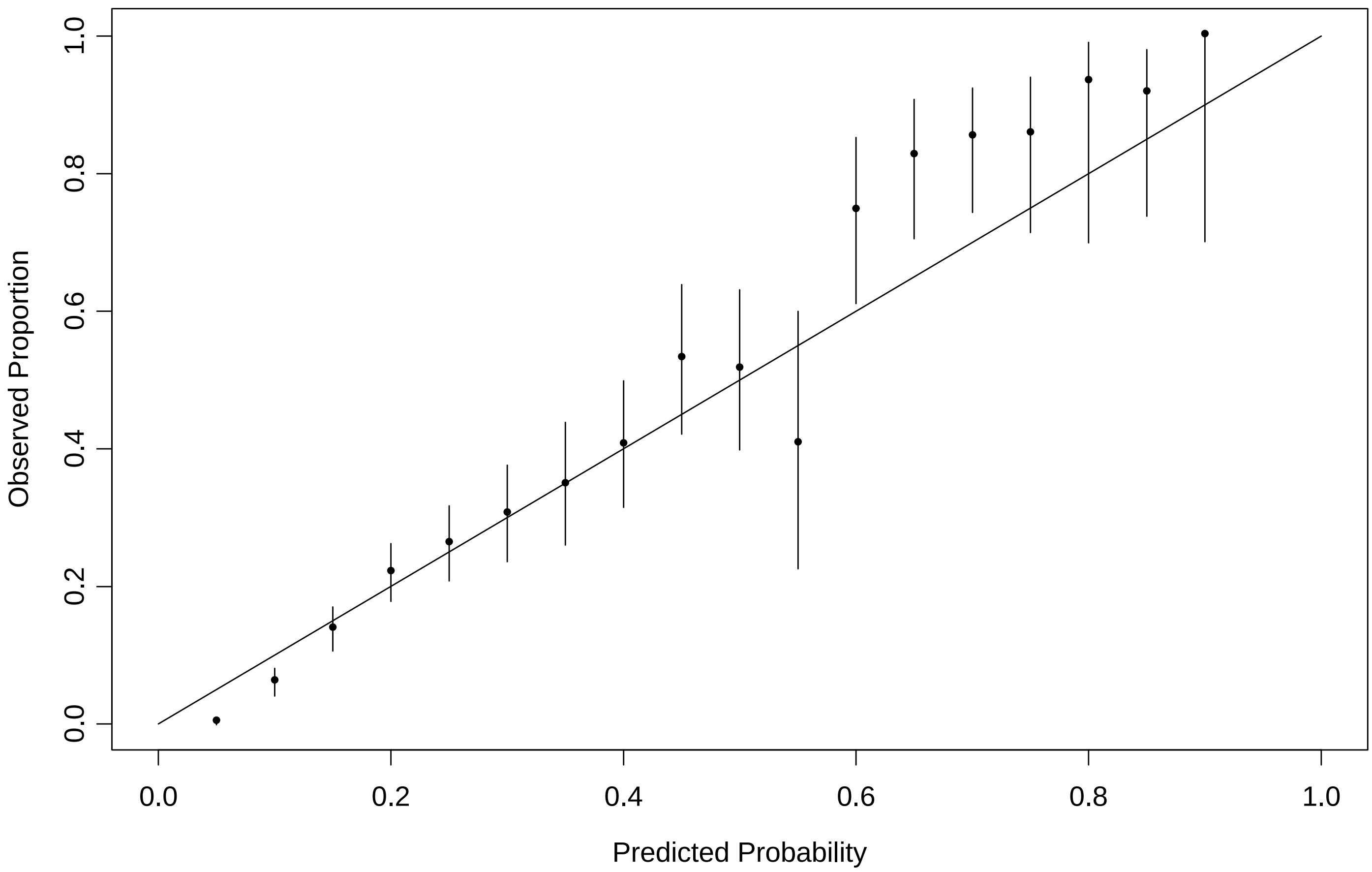


Figure 24: Relationship between the number of survey sites at which a given community is predicted to occur (derived by summing probabilities of occurrence predicted by a jackknifed model for each site) and the actual number of sites at which the community is recorded

Each symbol represents a different community.

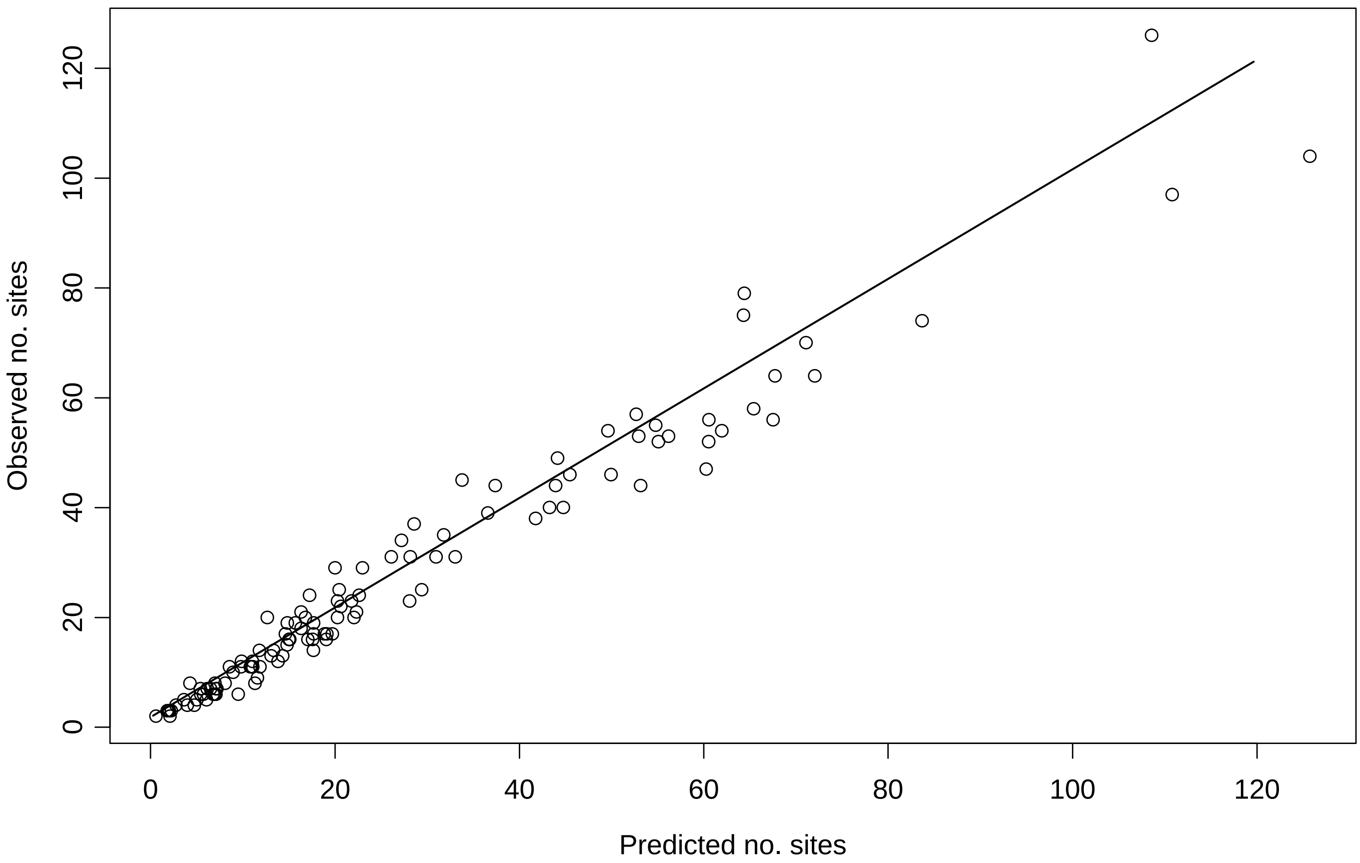


Figure 25: The 'sum of squares' (a measure of discrepancy between observed and predicted values) for each survey site.

Calculated by withholding this site from a jackknifed model based the remaining sites
A higher sum of squares value indicates a higher discrepancy between observed and predicted.

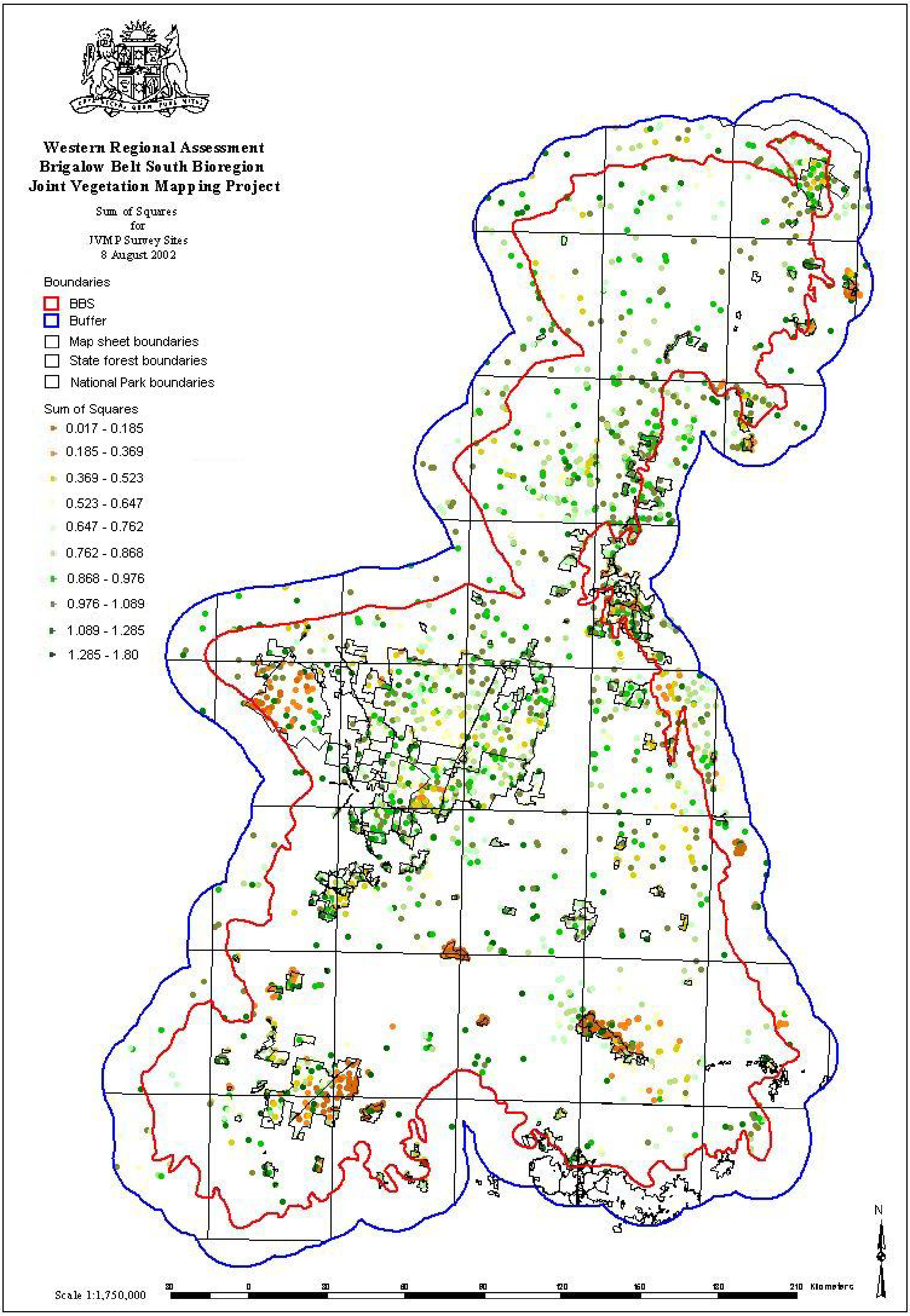
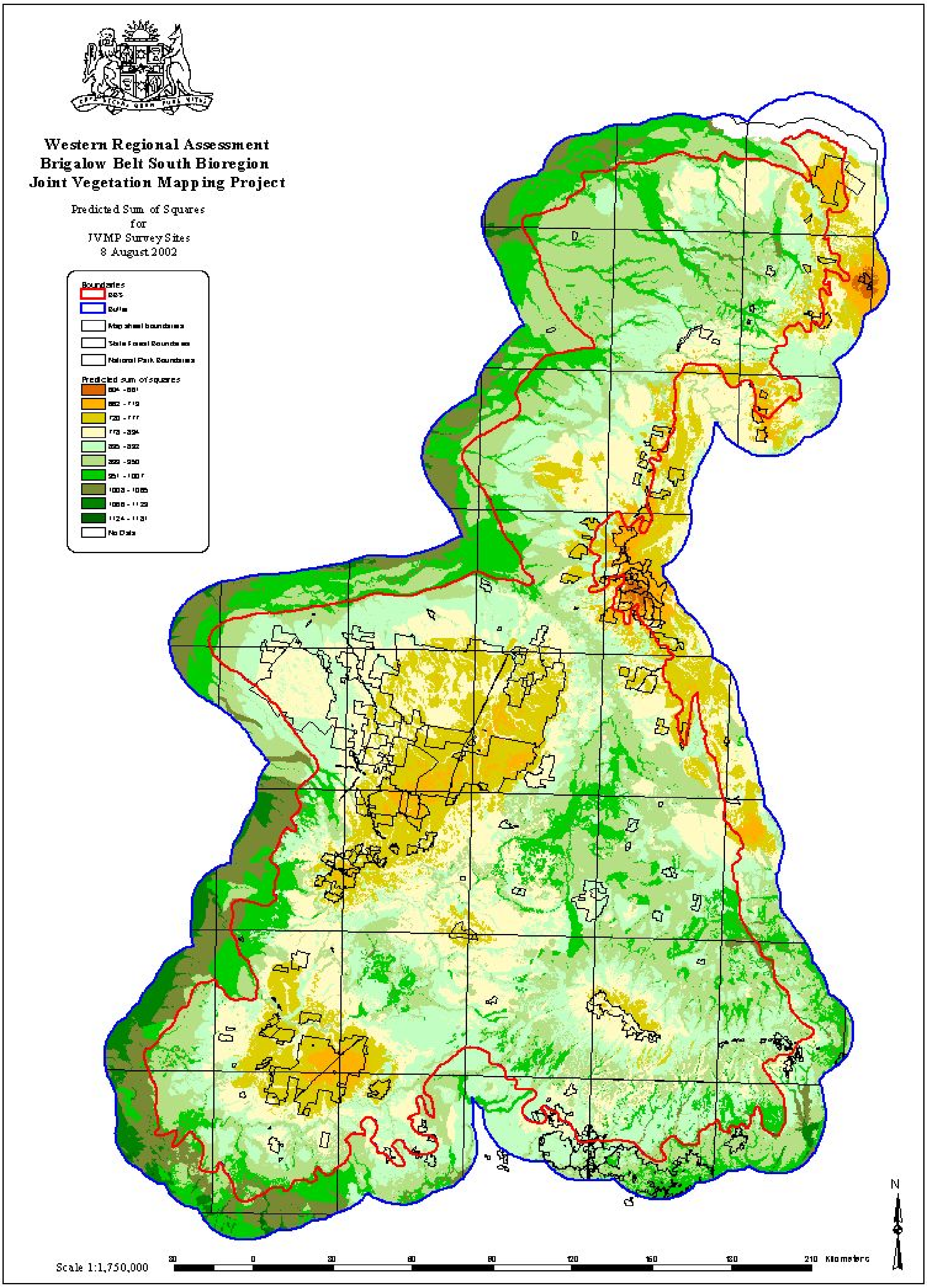


Figure 26: An approximate extrapolation of the sum of squares expected across all grid cells in the region

Based on a simple linear regression of the sum of squares calculated for each survey site against the density of other survey sites around this site (in terms of the transformed environmental / geographical space employed in the modelling). A higher sum of squares value indicates a higher discrepancy between observed and predicted, and therefore lower accuracy.



4.5.2 Predicted potential vegetation distribution

Probability surfaces

For each vegetation group the predicted potential distribution was expressed in terms of its gross area and net area.

The gross area is the sum of the 100 x 100 metres (one hectare) grid cells in which each vegetation group is predicted to occur. The gross area is therefore defined as the predicted area of each vegetation group within the BBS bioregion.

The net area is the sum of the probability of occurrence and the count of the grid cells for each probability. The net area is therefore the predicted area of occurrence for each vegetation group *within* the gross area. Appendix 8 provides area statistics for the predicted potential distribution of each vegetation group.

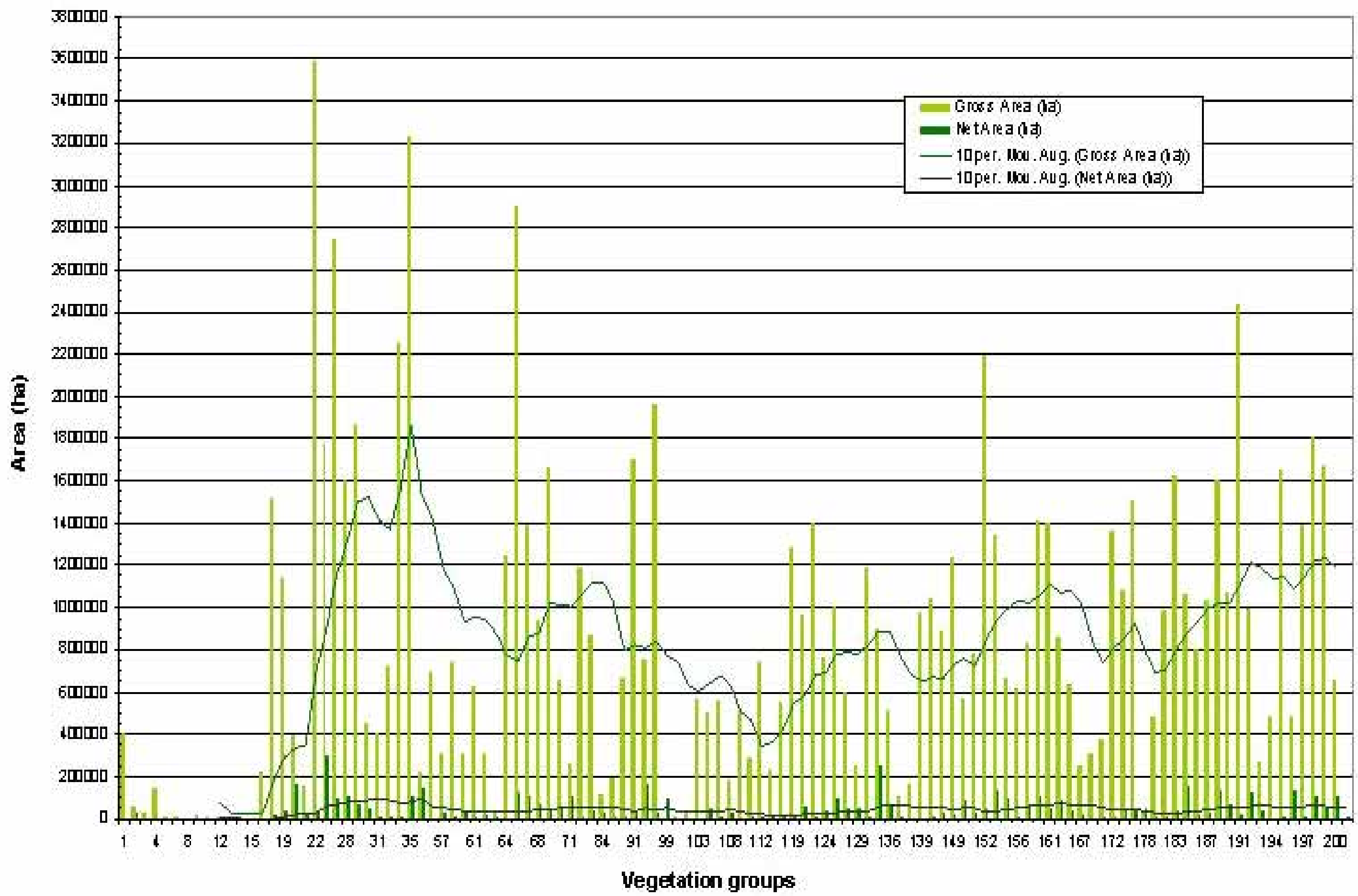
The net area-gross area ratio was calculated for each vegetation group along with statistics for the gross and net areas expressed as a percentage of the BBS bioregion area. The mean, maximum, minimum, range and standard deviation were also calculated for each vegetation group (Appendix 8). Figures are also provided for each vegetation group for the mean plus and minus one standard deviation. These indicate the positively skewed nature (ie the tail of the frequency distribution extends to the right) of the results, with the minimum probability for the means of all vegetation groups minus one standard deviation being one or zero. These findings were supported by the frequency distribution histograms, for each vegetation group, provided in Appendix 5. They suggest that the model does not discriminate well between vegetation groups at the gross level.

The total “net area” divided by the total “BBS bioregion area” is 98.17%. That is, the combined proportions of vegetation coverage predicted for every vegetation group accounts for 98.17% of the BBS bioregion (Appendix 8). The remaining 1.83% was accounted for in the model by rocky outcrops, waterways and wetlands.

Figure 27 illustrates the relationship between the predicted gross area for each vegetation group and the predicted net area. From the chart it can be seen that the relationship between gross area and net area is not close.

Probability surfaces of the potential vegetation distribution for each of the 115 vegetation groups are provided in Appendix 4 (a map accompanies this report). Appendix 5 contains histograms of the frequency distribution for each probability surface showing the relationship between the probability of occurrence and predicted area of occurrence.

Figure 27: Vegetation groups - predicted gross area and predicted net area distribution



From Appendix 8, vegetation group 22 (Pilliga cypress grass/herb woodland) had a predicted potential to have the largest gross area distribution within the BBS. Similarly, vegetation groups 35 (Eastern clay grassland), and 65 (Pilliga grassy cypress woodland), were predicted by the model to have the potential to occupy large areas of the BBS.

Conversely, vegetation groups which occur near Mt Kaputar (in the neighbouring bioregion) were predicted to have limited distributions within the BBS. For example vegetation group 15 (Kaputar grassy woodland) was modelled as having the predicted potential for a gross distribution of only 579 ha within the BBS. This is not to say that this vegetation group is rare or has a limited distribution but that its distribution within the BBS bioregion is limited to those environmental niches with which it is associated.

The predicted *net* area calculations also showed that group 22 had the potential to occupy the greatest area within the bioregion. Groups 130 (Goonoo ironbark heath woodland) and 19 (Coolah mixed woodland) had similarly been predicted to have the potential to occupy a large area of the BBS.

Vegetation groups with a low potential net area to gross area ratio are less likely to be the dominant vegetation group and may occupy niche locations within the landscape. These groups were predicted by the model as having less potential to successfully occupy large areas of the BBS bioregion. They generally have a low number of floristic survey sites associated with each vegetation group, and point to the inherent bias in the model due to an insufficient number of floristic survey sites across all vegetation groups.

4.5.3 Extant vegetation probability surfaces

Extant Probability surfaces

The composition of the extant vegetation was defined by the potential vegetation distribution model, which was constrained (or conditioned) by the composite API layer. The identification of the extant vegetation relied upon the modelled vegetation for its definition and the API and satellite image interpretation (where available) for its distribution.

Each of the 115 identified vegetation groups was represented in the extant vegetation layer. As with the predicted potential vegetation distribution model, the extant vegetation model for each vegetation group was derived from the constrained individual probability surfaces. Each probability surface was masked with the extant vegetation mask as described in section 3.13.2.

The composite API layer permitted identification of the overstorey vegetation with some degree of reliability. Due to the scale of the photography (1:50 000 scale) structural details, understorey floristics and land use information were not considered to be of a consistent, reliable and repeatable standard across all data sets and were therefore unable to be utilised by the JVMP. However, this information may be useful in providing local context when interpreting the JVMP outputs.

Appendix 9 details simple univariate statistics for the extant vegetation for each of the 115 probability surfaces. From Appendix 9, vegetation groups 22, 65, 35, 26 and 152 occupy the largest areas of gross extant vegetation. These five groups occupy a combined gross areal extent of 2 496 246 hectares or 93 % of the total extant gross vegetation. In contrast the net extant area occupied by these five vegetation groups is 400 641 hectares, or 16% of their combined predicted gross area.

Appendix 10 provides information on the ratios of gross predicted to gross extant area and net predicted to net extant area for each vegetation group. This allows for comparison between the current extant vegetation in hectares and the predicted potential vegetation distribution in hectares for each vegetation group. The ratio of these two measures provides an opportunity to assess the potential for revegetation for each vegetation group as well as acting as a guide for allocating revegetation priorities.

For example group 14 (Kaputar shrub woodland) has an extant net to predicted net area ratio of 97%. Appendix 10 shows that only 36 hectares of land within the BBS bioregion is predicted to meet the niche requirements of this vegetation group, that might be potentially available for revegetation. Appendix 10 does not consider landholder intent.

Using the net to predicted area ratios as a guide, group 14 may be deemed to have a lower priority for remedial works than group 49 (Riparian melaleuca woodland) which has a ratio of 27 % (the lowest ratio of all vegetation groups within the BBS). For group 49, a potential 1 847 hectares of land is available within the BBS bioregion with the abiotic variables required to support and meet its niche requirements. Again, this does not consider landholder intent.

Vegetation group 22 (Pilliga cypress grass/herb woodland) has the potential to cover up to 6% of the BBS. This group has a net extant to net potential area ratio of 43 %. Therefore, the potential exists that up to 172 229 hectares of suitable habitat could be targeted for land repair with a mixture of species from within that vegetation group.

Using Appendix 10, land managers could assess the relative status of each vegetation group by utilising the extant to predicted net area ratio. Appendix 10 could assist to decide where and how to allocate resources for land repair and revegetation projects. Table 13 provides a summary of the extant to predicted area ratios by ratio class.

Probability surfaces of the extant vegetation for each vegetation group are provided in Appendix 6.

TABLE 10: VEGETATION GROUPS EXPRESSED AS A PERCENTAGE OF NET AREA EXTANT/ PREDICTED RATIO CLASSES

Classes for net area — extant / predicted ratio	Count of vegetation groups within each class	Percent of Vegetation groups within ratio'd classes
<30%	1	< 1
31-40%	13	11
41-50%	29	25
51-60%	24	21
61-70%	17	15
71-80%	15	13
81-90%	8	7
91-100%	8	7
Totals	115	100

In total 2 739 814 hectares of extant vegetation was modelled and then mapped. Table 14 provides an overview of net area extant / predicted ratio classes. Where the modelled vegetation groups have a net area extant to predicted ratio of less than 30% they account for 689 hectares out of a predicted area of 2563 hectares. The predicted area equates to 0.05% of the total area of the BBS. Vegetation groups that fall into the 40-49% ratio class account for the largest proportion of extant vegetation within the BBS. From Table 14 880 455 hectares or 17% of the area of the BBS bioregion is represented by this class. Vegetation groups which fall within the 50-59% class account for 644 631 hectares or 12% of the area of the BBS.

Table 14 provides aggregated information by net area extant / predicted ratio classes about the areal extent for predicted gross area, extant gross area, predicted net area and extant net areas within the BBS.

TABLE 11: AREA BY AGGREGATED VEGETATION GROUP

Classes for net area — extant / predicted ratio	Predicted Gross Area (ha)	Extant Gross Area (ha)	Predicted Net Area (ha)	Extant Net Area (ha)
<30	226 432	66 608	2 536	689
30-39	11 028 795	3 933 968	602 364	195 160
40-49	39 937 821	16 755 235	2 020 057	880 455
50-59	22 264 831	10 229 175	1 186 110	644 631
60-69	12 194 699	6 485 321	618 640	396 306
70-79	7 952 266	4 515 201	4 618 577	341 102
80-89	3 271 638	2 159 753	258 181	217 091
90-100	118 109	105 349	4 540	4 264

5. DISCUSSION

The Joint Vegetation Mapping Project was established to gather information about the native vegetation of the Brigalow Belt South bioregion. Additionally the JVMP was tasked with the analysis of that information and further modelling to produce extant and predicted vegetation maps of the BBS. These products were key inputs to the Western Regional Assessment process, with a number of other key projects relying on the datasets produced either as an end product or during the process.

5.1.1 Vegetation groups within buffer areas

Of the 115 vegetation groups identified through the data analysis process a number of them fell almost exclusively within the 15 km buffer zone adopted around the BBS bioregion. Furthermore, a number of the vegetation groups which fell within the buffer were not widely distributed throughout the BBS. Rather than indicating rarity or restricted distribution, this observation served as a reminder that the buffer zone contained five neighbouring bioregions. It was expected that there would be some overlap of vegetation groups between bioregions, especially within such buffers.

5.1.2 Gap analysis bias and sampling bias

The gap analysis survey sites which had a woody vegetation mask applied were likely to be biased to the more extensively wooded parts of the landscape because the woody mask was applied after candidate sites were selected. Thus, the most extensively cleared groups were originally allocated a lower sampling intensity resulting in an inadequate definition of the vegetation groups present at those locations. Subsequently those vegetation groups of limited distribution were masked in the floristic analysis.

The JVMP floristic survey produced 1241 new survey sites across the BBS. Examination of the survey site locations at the completion of the survey stage revealed a bias in the sampling. Survey sites in the south of the BBS bioregion exhibited a bias towards crown lands. This was due to a number of factors including access constraints to privately managed property; resulting in a focus on the Crown estate by some botanists.

The initial analysis of the data suggested that areas to the immediate east and north of Dubbo, within the BBS, had not been sampled well enough to adequately define some vegetation groups known to occur in those areas.

Conversely, vegetation group 130 (Goonoo ironbark heath woodland) and group 21 (Coolah tops grass / herb forest) had the highest net area / gross area ratios of all vegetation groups. This suggested that these groups would be the dominant vegetation groups within their predicted areas. Both of these vegetation groups were well sampled with a large number of floristic survey sites per group, and so, the predicted distribution for each group was considered robust. However, due to the biases in the sampling regime groups 21 and 130 may have been be over modelled. The vegetation groups neighbouring groups 21 and 130 had fewer floristic survey sites per group by area. As a result, these groups were likely to be undermodelled.

5.1.3 Predictive accuracy of modelled vegetation

Vegetation groups with a low potential net area to gross area ratio were less likely to be the predicted dominant vegetation group. These groups may have occupied niche locations within the landscape. They were predicted by the model as having less potential to successfully occupy large areas of the BBS bioregion and generally had a low number of floristic survey sites associated with each vegetation group. As a result, the distinctiveness of these vegetation groups might not have been adequately recognised if they were masked in the floristic analysis.

Also of concern was a probable insufficient level of discrimination in the soil related abiotic variables used in the modelling process. Environmental differences in the landscape may not have been adequately represented in the variables used. This may have resulted in the models themselves being deficient in their ability to discriminate, even with adequate samples.

The predictive accuracy of modelled vegetation distributions such as those derived by the JVMP should ideally be evaluated using independent survey data, ie data collected at sites other than (and preferably well away from) those used in the original modelling (Pearce and Ferrier, 2000; Pearce, et al., 2001). Unfortunately, no independent survey data were readily available for use in the JVMP. While cross-validation goes some way towards affording independence between model-development and model-evaluation data, estimates of accuracy derived from such analysis are likely to still be optimistic, particularly if the locations of survey sites are biased or clumped, either geographically or environmentally.

Given the unavoidable bias in the JVMP surveys towards larger patches of extant woody vegetation, and the bias in distribution of this extant vegetation towards particular environments, evaluating predictive accuracy based on cross-validation of the survey data may provide a reasonable indication of the accuracy of modelling across extant vegetation. However, it is likely to overestimate the accuracy of predictions across poorly sampled areas such as privately managed land. Unbiased estimation of predictive accuracy across these areas will require further survey effort to collect appropriate independent evaluation data.

Appendix 8 indicates that of the 115 vegetation groups, only 11 had a predicted net area to gross area ratio greater than 10% and only one vegetation group had a net area to gross area ratio greater then 20%. If the ratio between the predicted net area and predicted gross area is high then there can be greater confidence that the model will more accurately predict where each vegetation group will occur. That is, it will discriminate well between vegetation groups.

In this instance, the predicted net area-gross area ratio for the vegetation groups is very low. More than 90% of the modelled vegetation groups have ratios of less than 10%, and more than 64% of the vegetation groups have ratios of less than 5%.

The sum of the predicted gross area for all vegetation groups was 18.5 times the area of the BBS bioregion. This has implications about the ability of the model to predict vegetation group type and location. While the model may be predicting the distribution of some vegetation groups well, it is not able to predict what vegetation group will occur within any one patch of land within the bioregion with a high degree of confidence.

This is especially problematic with vegetation groups which are floristically and environmentally closely related and there is an equal chance of several groups occurring in one area. A direct outcome of this is that the modelled vegetation groups may be poor predictors of vegetation distribution across the bioregion and should not be used for property scale planning or detailed mapping.

These issues are not new or restricted to the JVMP. Strategic landscape level mapping will result in loss of detail at the local scale. This reduces the utility of the model as a tool for predicting the vegetation groups likely to occur within a particular parcel of land.

The extant vegetation within the BBS bioregion accounted for 52% of the area of the bioregion. Undoubtedly some of the areas modelled as containing extant vegetation will not contain the full suite of species for the particular vegetation group predicted to occur within that area. This is not especially problematic for landscape level planning purposes as the variation in the vegetation condition and composition was captured at the survey site and API polygon level at scale favourable to bioregional planning.

6. RECOMMENDATIONS - IMPROVING FUTURE REGIONAL VEGETATION MAPPING

Issue 1: The key limitation of the JVMP was the limited number of floristic samples. The combined floristic surveys utilised by the JVMP provided 3 166 survey sites for analysis. These included a mixture of 20 metres x 20, and 20 metres x 50 sites. In total, approximately 300 hectares of the BBS bioregion was sampled out of a total area of 5 250 434 ha. That is, on average there was only one survey site per 1 658 hectares, or, put another way, each hectare sampled represented 17 500 hectares of vegetation.

Sampling stratification was also an important issue due to private land access issues.

These sampling limitations affected the PATN analysis, the construction of the environmental space, the modelled vegetation, and the utility of the final product. The author suggests that, where possible, sampling density should be about three times greater for this scale of mapping.

Recommendation 1: That, if time and resources allow, a minimum standard be adopted for vegetation survey for regional assessments that would result in a minimum survey effort of one site per 500 hectares.

Issue 2: A second limitation of the JVMP is that it was not possible to fund bioregion wide aerial photography interpretation. This resulted in the use of targeted API, of the woody vegetation for specified map sheets, and pre-existing API. The scale of the photography used by the JVMP was 1:50 000. The use of this scale of photography for this scale of vegetation mapping is not recommended as detail is lost and the cost of interpretation increases compared with 1:10 000 and 1:15 000 scale photography.

With a large number of API data sets used by the JVMP, inconsistent quality control and interpretation standards resulted in a lowest common denominator approach being adopted when the API was used. This resulted in a loss of detail for the JVMP in the interests of having a consistent and repeatable final data set.

Recommendation 2: That if API is to be utilised for bioregional assessment, RACAC to adopt a scale of 1:15 000 as the preferred API standard. Alternatively, if such scale photography is unavailable then API should be enhanced with or replaced with alternative remote sensing technologies at a bioregion wide scale for the Regional Assessment process.

Issue 3: Satellite imagery was utilised during the project to help overcome the limitations imposed from patchy availability of fine resolution aerial photographs. Satellite imagery was delivered in a timely manner and allowed the differentiation of forest, woodland, grassland and urban interfaces. For the purposes of defining an extant vegetation layer from the modelled data, the satellite image interpretation proved itself invaluable.

Satellite imagery is becoming increasingly useful in land use planning, and the resolution of some products allows for detailed mapping of natural resource features, although ground survey to validate such mapping is still required. When field data is combined with digital mapping techniques, satellite data can improve the accuracy of vegetation modelling, and can be a useful monitoring tool, at competitive costs.

However, for detailed vegetation structure, understorey, age class and condition information, satellite imagery may not be at a stage where it could match the quality of fine scale API and ground survey; although this is a nearing possibility as the price of high resolution satellite imagery becomes more affordable, as computer power increases, and as expertise expands from API mapping to using other forms of remote sensing. Recent release satellite imagery is available at a scale equivalent to 1:10 000 photography. Analysis techniques allow repeatable and consistent method to be utilised.

Recommendation 3: The use of high resolution satellite imagery should be increasingly considered in future regional assessments, especially if recent and fine scale aerial photographs are not available. Good ground survey to validate satellite derived mapping is still required.

Issue 4: A limitation of the modelling process adopted for the JVMP was the inability to define which vegetation groups occurred together. For example if vegetation group A has a gross area of 100 000 hectares and a net area of 15 000 hectares then the end user may want to know what other vegetation groups occur within the gross area for that vegetation group and all other vegetation groups.

Where the probabilities of occurrence are low for the majority of the vegetation groups then these relationships are important in determining if, for instance, the groups require merging.

Where multiple vegetation groups are predicted to occur in a single pixel, a tool that allows easy spatial identification of those groups would be valuable. To identify spatially which vegetation groups are predicted to occur over a group of neighbouring pixels would also be valuable. Currently each probability surface needs to be interrogated separately to determine which vegetation groups are predicted to occur over any group of neighbouring pixels. In this project, it was difficult for JVMP agency staff to easily and reliably interrogate the 118 probability surfaces for understanding the spatial distribution of the predicted vegetation groups.

Recommendation 4: That the assessments include the development of a tool for use during the projects to run queries on the vegetation groups that occur together, and across neighbouring pixels, for regional planning.

7. REFERENCES

Bedward, M., Keith, D.A. & Pressey, R.L, 1992, 'Homogeneity analysis: assessing the utility of classifications and maps of natural resources', *Australian Journal of Ecology*, 17, 133-139.

Belbin, L. & McDonald, C. 1993, 'Comparing three classification strategies for use in ecology'. *Journal of Vegetation Science* 4, 341-348.

Belbin, L. 1995, PATN, Pattern Analysis Package, CSIRO, Canberra, 234 pp.

Bray, J. R. & Curtis, J. T. 1957, 'An ordination of the upland forest communities of southern Wisconsin', Ecological Monographs, 27: 325-349.

Duivenvoorden, J. F., Svenning, J.C. & Wright, S. J. 2002, 'Beta diversity in tropical forests', *Science* 295: 636-637.

DLWC, 2002, Soil Landscape Reconnaissance Mapping Project, Brigalow Belt South Bioregion Western Regional Assessment Stage 2, NSW Department of Land and Water Conservation, Resource and Conservation Assessment Council, Sydney.

Efron, B. & Tibshirani, R. 1993, 'An introduction to the bootstrap'. Chapman and Hall, New York.

Faith, D.P. Minchin, P.R. & Belbin, L. 1987, 'Compositional dissimilarity as a robust measure of ecological distance', *Vegetation* 69, 57-66.

Faith, D. P. & Ferrier, S.2002, 'Linking beta diversity, environmental variation, and biodiversity assessment', *Science* 296 [Online http://www.sciencemag.org/cgi/eletters/295/5555/636#504], 22 July 2002.

Ferrier, S. 2002, 'Mapping spatial pattern in biodiversity for regional conservation planning: where to from here?', *Systematic Biology* 51: 331-363.

Ferrier, S. Drielsma, Manion, M. & Watson, G. 2002, 'Extended Statistical Approaches to Modelling Spatial Pattern in Biodiversity in North-East New South Wales: II', 'Community-level modelling, Biodiversity and Conservation' 11: in press.

Ferrier, S. Gray, M. R. Cassis, G. A. & Wilkie L. 1999a, 'Spatial Turnover in Species Composition of Ground-Dwelling Arthropods, Vertebrates and Vascular Plants in North-East New South Wales: Implications for Selection of Forest Reserves. Pages 68-76 in The Other 99%, The Conservation and Biodiversity of Invertebrates (W. Ponder and D. Lunney, eds.). Royal Zoological Society of New South Wales, Sydney.

Ferrier, S., Pearce, J., Drielsma, M., Watson, G., Manion, G., Whish, G., & Raaphorst S. 1999b, 'Evaluation and refinement of techniques for modelling distributions of species, communities and assemblages', Unpublished report prepared for Environment Australia, New South Wales National Parks and Wildlife Service, Sydney.

Houlder,D.J., Hutchinson,M.F., Nix,H.A. & McMahon,J.P. 1999, ANUCLIM User Guide, Version 5.0. Centre for Resource and Environmental Studies, Australian National University, Canberra.

Jorgensen, S.E. 1994, 'Fundamentals of Ecological Modelling' (2nd ed), Elsevier, Amsterdam.

Keith, D. A. & Bedward, M. 1999, 'Native vegetation of the South East Forests region, Eden, New South Wales'. *Cunninghamia* 6: 1-218.

Legendre, P., Lapointe, F.-J. & Casgrain P. 1994, 'Modeling brain evolution from behaviour: a permutational regression approach', *Evolution* 48: 1487-1499.

Lowe, D. G. 1995, 'Similarity metric learning for a variable-kernel classifier', *Neural Computation* 7: 72-85.

Lunt, I. 1997, Effects of long-term vegetation management on remnant grassy forests and anthropogenic native grasslands in south-eastern Australia. *Biological Conservation* 81, pp287-297.

McCullagh, P. & Nelder, J. A, 1989, 'Generalised Linear Models', 2nd edition, Chapman and Hall, London.

Moore, D. M., Lees, B. G. & Davey, S. M. 1991, 'A new method for predicting vegetation distributions using decision tree analysis in a geographic information system'. *Environmental Management* 15: 59-71.

Murray-Darling Basin Commission. 2000, Basin-in-a-Box, Murray Darling Basin CD set, produced and distributed under licence for the Murray-Darling Basin Commission by Georeality Pty Ltd.

NPWS. 2002, 'Yet Another Vegetation Survey Database', NPWS Vegetation Survey Database, NSW National Parks and Wildlife Service, Hurstville.

Oliver, I., Smith, P.L., Lunt, I. & Parkes, D. 2002, 'Pre-1750 vegetation, naturalness and vegetation condition: What are the implications for biodiversity?' *Ecological Management and Restoration* Vol 3 No3 pp177-179.

Pearce, J. & Ferrier, S. 2000, 'Evaluating the predictive performance of habitat models developed using logistic regression'. *Ecological Modelling* 133: 225-245.

Pearce, J. Ferrier, S. & Scotts D. 2001, 'An evaluation of the predictive performance of distributional models for flora and fauna in north-east NSW', *Journal of Environmental Management* 62: 171-184.

Poulin, R. & Morand, S. 1999, 'Geographical distances and the similarity among parasite communities of conspecific host populations', *Parasitology* 119: 369-374.

Prescott, J. A. 1948, 'A climatic index for the leaching factor in soil formation', *Journal of Soil Science.* 1, 9-19.

RACAC. 1999,'Modeling areas of habitat significance for vertebrate fauna and vascular flora in NE NSW'. Unpublished report to the Resource and Conservation Assessment Council

Sivertsen, D. & Smith, P.L, 2001, ‘Guidelines for Mapping Native Vegetation - Version 2.1’, Unpublished draft, DLWC Centre for Natural Resources, Parramatta.

Smith, P.L., 2000, ‘An overview of existing pre-1750 modelling techniques used in NSW and a proposed alternative strategy for Inland NSW’. *Report to the National Land and Water Resources Audit*. NSW Department of Land and Water Conservation, Parramatta.

Stehman, S. V. & Czaplewski, R. L. 1998, ‘Design and analysis for thematic map accuracy assessment: fundamental principles’, *Remote Sensing of Environment* 64: 331-344.

Strahler, A. H. 1980, ‘The use of prior probabilities in maximum likelihood classification of remotely sensed data’, *Remote Sensing of Environment* 10: 1135-1163.

Winsberg, S. & De Soete, G. 1997, ‘Multidimensional scaling with constrained dimensions: CONSCAL’. *British Journal of Mathematical and Statistical Psychology* 50: 55-72.

8. APPENDICES

(MAPS AND CD-ROM)

1. Floristic survey data capture sheet
2. Aerial photography mapping pathway
3. Vegetation community descriptions
4. Predicted potential vegetation distribution — probability surfaces
   - MAP FOLDER
5. Predicted potential vegetation distribution — frequency distributions.
6. Extant vegetation distribution — probability surfaces
   MAP FOLDER.
7. Dendrogram of group associations for final groups
8. Potential vegetation group —predicted area statistics
9. Extant vegetation group — predicted area statistics
10. Predicted vegetation distribution ratios
11. Survey Gap Analysis Tool