acera
Australian Centre of
Excellence for Risk Analysis

Report Cover Page

ACERA Project		
0706		
Title		
Evaluating vegetation condition measures for cost effective biodiversity investment planning		
Author(s) / Address (es)		
Emma Gorrod[1,2] and David A. Keith[2] [1] School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia. [2] NSW Department of Environment and Climate Change, PO Box 1967, Hurstville, NSW 2220, Australia.		
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)		
Manuscript (final report)		
Summary		
This study was motivated by the call from Auditor General to provide reliable feedback on NHT investments. This study summarises the magnitude of observer error in measures of vegetation condition developed by the states. Errors are random and appreciable, and the study indicates the most important areas where improved field protocols will substantially improve the quality of monitoring and reporting.		
ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	DAFF Endorsement: ( ) Yes ( ) No	Date:

Evaluating vegetation condition measures for cost-effective biodiversity investment planning;
ACERA Project No.0706

Kenton Lawson
ABARE

Observer variation in field assessments of vegetation condition: Implications for biodiversity conservation

Emma Gorrod[1,2] and David A. Keith[2]

[1] School of Botany, University of Melbourne, Parkville, Victoria 3010, Australia.
[2] NSW Department of Environment and Climate Change, Hurstville, NSW 2220, Australia.

Manuscript (final report)

July 2008

Acknowledgements

This report is a product of the Australian Centre of Excellence for Risk Analysis (ACERA). In preparing this report, the authors acknowledge the financial and other support provided by the Department of Agriculture, Fisheries and Forestry (DAFF), the University of Melbourne, Australian Mathematical Sciences Institute (AMSI) and Australian Research Centre for Urban Ecology (ARCUE).

Many thanks to the observers that participated in the field trial, and their employers: NSW Department of Environment and Climate Change, Victorian Department of Environment and Sustainability and Biosis Research Pty. Ltd. The manuscript was greatly improved by comments from Ian Oliver and another reviewer. Thanks also to Mark Burgman and Murray Ellis for editorial assistance. EG was supported by an Australian Postgraduate Award.

Disclaimer

Table of contents

List of Tables

List of Figures

Executive Summary

Assessments of vegetation condition are often used to inform land management and planning decisions for biodiversity conservation, such as allocation of incentive funding and determination of offset actions to compensate for biodiversity losses. Uncertainty in assessments of vegetation condition may lead to poor decisions or unexpected outcomes, including the loss of biodiversity. This study investigates uncertainty in assessments of vegetation condition due to observer error in field estimates of vegetation attributes. Ten observers conducted vegetation condition assessments using two assessment protocols (BioMetric and Habitat Hectares) on 20 sites in a grassy woodland community. Observers' estimates varied substantially across multiple scoring categories for all vegetation attributes on almost all sites. Observers generally agreed on the total scores and ranks of highly degraded (pasture) sites, but were less consistent on other sites. Across all sites, the average coefficient of variation was 18% for BioMetric and 15% for Habitat Hectares, and the maximum was 60%. All observers estimated vegetation condition scores that were substantially different from the group mean on at least some sites. The results indicate that uncertainty in field estimates of site attributes may cause vegetation condition to be under- or over-estimated on all but highly degraded sites. The primary cause of observer variation in total vegetation condition scores was random error in raw estimates of vegetation attributes, rather than differences in the index structures or sampling methods. It is recommended that: research is undertaken into methods for reducing observer error; field observers estimate uncertainty around point estimates of vegetation condition; the sensitivity of index scoring structures to observer error is reviewed; and that decision makers explicitly incorporate uncertainty into the decision making processes and aim for outcomes that are robust to this uncertainty.

Introduction

Quantitative estimates of vegetation attributes are frequently required to inform land management decisions for biodiversity conservation. Relevant decisions include the allocation of incentive funding to manage private land for biodiversity conservation (USDA 2003; Oliver et al. 2005) and determination of offset actions to compensate for unavoidable biodiversity losses (ten Kate et al. 2004; DEC 2005a; DSE 2006). Decisions that fail to consider uncertainty in field assessments of vegetation attributes may lead to unexpected outcomes, including loss of biodiversity. Observer error in field estimates of vegetation attributes may be an important cause of uncertainty in land management decisions for biodiversity conservation (for others see Gorrod et al. *in review*). Yet decisions in the context of biodiversity conservation are usually made as though the values of vegetation attributes are known without error.

Observer error may occur due to inaccurate estimation of a quantity (measurement error) or failure to correctly identify or interpret the feature to be estimated (identification error). Measurement error arises from observational techniques or instrument error, and varies randomly about the true value (Regan et al. 2002). Identification errors may occur as a result of linguistic uncertainty (imprecise language) in the protocol's survey methodology. Identification error may occur, for example, if an observer estimates the length of all logs rather than only those logs that exceed a specified diameter, or assesses projective foliage cover rather than canopy cover (e.g. Gray and Azuma 2005). Some observers may systematically over- or under-estimate attribute values. Systematic bias and identification error may confound measurement error such that variation across estimates of multiple observers does not centre on the true value of the parameter.

Observer error is reflected in imprecision amongst the estimates of multiple observers. Research into precision of multiple observers estimating vegetation cover has reported coefficients of variation between 10 and 200% (Sykes et al. 1983; van Hees and Mead 2000; Klimes 2003). As true values of vegetation attributes are seldom known, it is difficult to estimate accuracy of observers' estimates, particularly for cover estimates. Underestimation of counts (detectability) has been documented for attributes such as hollow bearing trees (Harper et al. 2004) and plant species richness (e.g. Hellmann & Fowler 1990; Ringvall et al. 2005).

Observer error in field estimates of vegetation attributes may be exacerbated by the mathematical structure of multivariate indices used to quantify biodiversity value at the site scale. Conversion of raw attribute estimates into categorical scores may exacerbate error, for instance, if the placement of scoring thresholds is such that small errors in raw estimates cause changes in class membership. Scoring errors may be compounded if combined multiplicatively.

Different protocols for quantifying biodiversity value at the site scale, including field survey techniques and multivariate index structures, are developed by different jurisdictions to meet their policy requirements. The Environmental Benefits Index, for example, is used to rank sites for incentive funding in the Conservation Reserve Program in the United States (USDA 2003). In Australia, vegetation condition indices are used to allocate incentive funding and determine offset requirements. The BioMetric (Gibbons et al. 2005) and Habitat Hectares (Parkes et al. 2003; DSE 2004) protocols are used in New South Wales and Victoria, respectively. Both protocols require field assessments of vegetation structural and composition attributes (listed in Table 1), which are selected on the basis that they are surrogates for habitat features required by indigenous species or indicators of ecological processes. Both protocols allocate scores for each site attribute relative to reference (or benchmark) values. Scores for individual attributes are combined to yield a total vegetation condition score that represents the similarity of the site to a benchmark stand of the same vegetation community.

We aimed to address the following questions:

(1) What is the magnitude of variation among assessors in their assessment of attributes contained within contemporary vegetation condition assessment tools?
(2) What is the impact of this variation on vegetation condition metrics generated by two contemporary vegetation condition assessment tools?
(3) What are the implications for biodiversity conservation of the results for (1) and (2)?

Methodology

Protocols for assessing vegetation condition

The BioMetric and Habitat Hectares indices contain similar sets of vegetation attributes, which are estimated, weighted and combined using different methods (Oliver et al. 2007) (Table 1). Though both use relatively broad scoring intervals for each attribute, in part to accommodate a margin of observer error, the size and number of scoring intervals differs (see Table 3 in Gibbons et al. 2005 and Appendix 8 in DSE 2004). The total BioMetric score is calculated by combining the individual attribute scores using Equation 1 to yield a score out of 100, whereas Habitat Hectares total score is the sum of attribute scores with a maximum score of 75. In this study, Habitat Hectares scores have been standardised to a maximum value of 100.

$$\text{BioMetric score} = \frac{\left(\sum_{v=a}^{j}(s_v \cdot w_v)\right) + 5\left((s_a \cdot s_g) + (s_b \cdot s_i) + (s_h \cdot s_j) + (s_c \cdot s_k)\right) \times 100}{480} \qquad (1)$$

S_v is the score for attribute a-j (see Table 1), w_v is the weight of the attribute and S_k is the average of scores for attributes d, e and f. Selected pairs of attributes are combined as products to reflect ecological relationships (S. Briggs, pers. comm., 2007). The multiplication of scores for *number of native plant species* and *lack of exotic cover*, for example, implies that they are not directly substitutable and co-occurrence of both attributes would substantially improve the value of a site (Gibbons and Freudenberger 2006). The denominator, 480, is the maximum possible score for a community in which values for all ten attributes fall within the benchmark.

Table 1. BioMetric and Habitat Hectares component attributes, weightings and Cumberland Plain Woodland benchmarks.

BioMetric				**Habitat Hectares**		
Attribute		**Benchmark**	**Weight**	**Pairs of attributes**	**Benchmark**	**Weight**
H	Number of hollow bearing trees	≥1 tree	30	Number of large trees per hectare	>15 trees with dbh ≥50cm	10
				Large tree canopy health	>70%	
B	Native over-storey cover	19-24%	5	Tree canopy cover	7-22%	5
				Tree canopy health	>70%	
A	Native plant species richness	>29	20	Diversity & cover of understorey lifeforms	≥9 lifeforms present	25
C	Native mid-storey cover	20-30%	10			
D	Native ground cover (grasses)	23-31%	5	Proportion of understorey lifeforms that are substantially modified	0 lifeforms modified	
E	Native ground cover (shrubs)	0-5%	5			
F	Native ground cover (other)	12-20%	5			
g	Cover of weeds	0-5%	5	Cover of weeds	0-5%	15
				Proportion of weed species that are considered high threat	0	
i	Proportion of overstorey species regenerating	100%	10	Total number of woody species	≥5	10
				Proportion of woody species recruiting	>70%	
j	Total length of logs	≥5 m of logs ≥10cm diameter	5	Total length of logs	≥7.5 m of logs of ≥10cm diameter	5
				Proportion of logs that are large	≥2.5 m of logs of ≥25cm diameter	
				Litter cover	5-15%	5
				Dominance of native/exotic litter	Native litter	
	Total		**100**	**Total**		**75**

BioMetric assessments are conducted in 20 x 50 m plots whereas Habitat Hectares assessments are conducted within an area of unlimited size containing relatively homogeneous vegetation. Cover attributes are visually estimated for Habitat Hectares, whereas observers may choose to either visually estimate or use point count techniques for estimating native shrub cover, native ground cover and total exotic cover for BioMetric. Both BioMetric and Habitat Hectares protocols supply operational manuals (Gibbons et al. 2005 and DSE 2004 respectively), which contain different schematics for assisting observers to make cover estimates. The BioMetric protocol requires the observer to specify a point estimate which is then converted into a score, whereas the Habitat Hectares protocol requires the observer only to select a scoring category. The landscape context components of the indices were not addressed in this study.

Observers

Ten observers were selected to represent a sample of all possible observers that may carry out vegetation condition assessments. The observers all had relevant tertiary qualifications and previous experience conducting vegetation surveys, and included the authors (Observers A and B) (Table 2). General experience ranged from approximately two years experience as an environmental consultant (Observer J) through to 26 years experience as a plant ecologist (Observer B). Observers E and F were relatively more experienced in the use of the Habitat Hectares and BioMetric methodologies respectively.

Table 2. Characteristics of observers in the field trial.

Code	Current occupation	Years of general vegetation survey experience	Familiarity with Cumberland Plain Woodland[1]	Familiarity with survey methods (BioMetric/Habitat Hectares) [2]
A	PhD candidate	5	Mod	low/mod
B	Principle Research Scientist, DECC[3]	26	High	low/low
C	Vegetation dynamics Research Officer, DECC	4	Low	low/low
D	Vegetation dynamics Project Officer, DECC	7	Low	low/low
E	Native Vegetation Project Officer, DSE[4]	4.5	Nil	low/high
F	Information and assessment Officer, DECC	10	Mod	mod/low
G	Vegetation dynamics Research Officer, DECC	8	Mod	low/low
H	Environmental consultant (flora)	5	High	low/low
I	Environmental consultant (flora)	8	Mod	low/low
J	Environmental consultant (flora, fauna & hydrology)	2	Mod	low/low

1: Self rated familiarity with Cumberland Plain Woodland. 2: Relative familiarity with (i.e. field experience using) BioMetric and Habitat Hectares. 3: New South Wales Department of Environment and Climate Change. 4: Victorian Department of Sustainability and Environment.

Field trial

All observers independently conducted vegetation condition assessments using BioMetric and Habitat Hectares protocols on 20 sites in Cumberland Plain Woodland (CPW), west of Sydney metropolitan area, Australia. CPW is a grassy woodland community that occurs on shale derived soils and is listed as an Endangered Ecological Community on state and commonwealth legislation (ESSS 2000; NPWS 2004). Sites were selected to represent a spectrum of structural and compositional variants of CPW, in which each of the canopy, shrub and ground strata were a) either structurally intact or modified and b) dominated by either native or exotic plant species (Table 3). All sites were located in reserves at the time of survey, but had been subject to a range of management histories including grazing, clearing, logging, fertiliser application, planting, weed control and conservation. Each site consisted of an area of relatively homogeneous vegetation and ranged in size from approximately 0.5 to 3 hectares. The available BioMetric benchmark for CPW was unmodified from the published version (DEC 2006). The benchmark for Habitat Hectares was composed on the basis of expert opinion (DK), and available floristic and structural survey data (Tozer 2003) (Table 1).

Table 3. Structural and compositional characteristics of sites 1 to 20, and their locations (Western Sydney Regional Parklands (WSRP), Planning NSW, Prospect Reservoir and Scheyville National Park).

Tree canopy	**Absent**		**Present (Planted)**		**Present (Non planted)**			
Shrub and ground layer	Absent	Dense native	Absent or sparse	Dense native	Absent or sparse	Dense exotic	Dense native	Open native
WSRP	10, 15		1, 3, 8, 11	2		5, 12		7, 20
PlanningNSW					4			
Prospect		19			16, 17		18	
Scheyville		9				6	13	14
Total	**2**	**2**	**4**	**1**	**3**	**3**	**2**	**3**

At each site, a floristic survey in which the occurrence of all native and exotic plant species were recorded within a randomly located 20 x 20 m quadrat was conducted by Observer A and at least one other observer. Due to observers' time constraints, the floristic data were then available to all observers conducting vegetation condition assessments. This reduced variation between observers for estimates of lifeform richness that may have been attributable to differences in plant identification skills. The native plant species richness attribute of the BioMetric method was estimated for all observers from the common floristic data set.

Assessments of vegetation condition were conducted by up to five independent observers on the same day between January and November 2006. The 20m x 50m quadrat used in the BioMetric method, and the boundary of the 'zone of relatively homogeneous vegetation' used in the Habitat Hectares method, was positioned in the same location for all observers, and encompassed the 20m x 20m floristic plot. Observers B, C and D conducted most assessments concurrently between January and October 2006; Observers E, F and G conducted all assessments concurrently over two weeks in May 2006; and Observers H, I and J conducted most assessments concurrently between May and October 2006. Weather conditions were reasonably stable throughout the duration of the surveys, and the vegetation attributes of CPW are not prone to significant seasonal fluctuations. The order in which the sites were visited was randomised amongst groups and the order in which each protocol was used was randomized amongst observers and sites. In accordance with the field assessment protocols, observers recorded point estimates for BioMetric attributes and selected scoring categories for Habitat Hectares attributes.

Observer A was trained by an officer from the Victorian Department of Sustainability and Environment in conducting field assessments using the Habitat Hectares method. The first time that other observers conducted assessments, Observer A trained them in the use of each assessment protocol, using the supporting documentation and guidelines provided (DSE 2004 and Gibbons et al. 2005). The documentation was available at all times for observers to consult for the remainder of the assessments. No attempt was made to calibrate estimates of percent cover amongst observers. Observers were not permitted to discuss their estimates with others, including the interpretation of terms beyond definitions provided in the protocols.

Data analysis

We analysed the magnitude of variation amongst observers' estimates of the attributes contained within BioMetric by plotting the range of observer estimates, and calculating Coefficients of Variation (CV), on each site. CV is calculated by dividing the standard deviation by the mean (of ten observers' estimates) and provides a unitless measure of the variation that is comparable across different types of vegetation attributes. CV is sensitive to small changes in the mean when the mean value is near zero. Variation was not calculated in this way for Habitat Hectares attributes as point estimates were not specified in field assessments.

The spread of observer estimates of vegetation attributes across BioMetric scoring categories was plotted, as was the spread of scores for Habitat Hectares attributes. CV of total scores was calculated to determine whether the magnitude of variation amongst observers was related to site types. The total score of each site averaged across observers was plotted for BioMetric and Habitat Hectares, and spearman rank correlation calculated, to determine how well the two metrics correlated. Spearman rank correlations were calculated for total score estimates between all pairs of observers, to test the direction and strength of relationships amongst observers.

Results

Variation in attribute estimates amongst observers

For all attributes recorded in the BioMetric assessment on each site, we calculated the mean and CV of the ten observers' estimates. For all attributes, CV declined with increasing mean, and for any given mean the magnitude of CV tended to vary more across sites than attributes. However, CV of *native ground cover (shrubs)* tended to be comparatively higher, and *native overstorey cover* comparatively lower, than other attributes (Figure 1). For means of 5 and under CVs ranged from 40% to 300%, and CV was not less than 20% for means up to 75 (Figure 1).

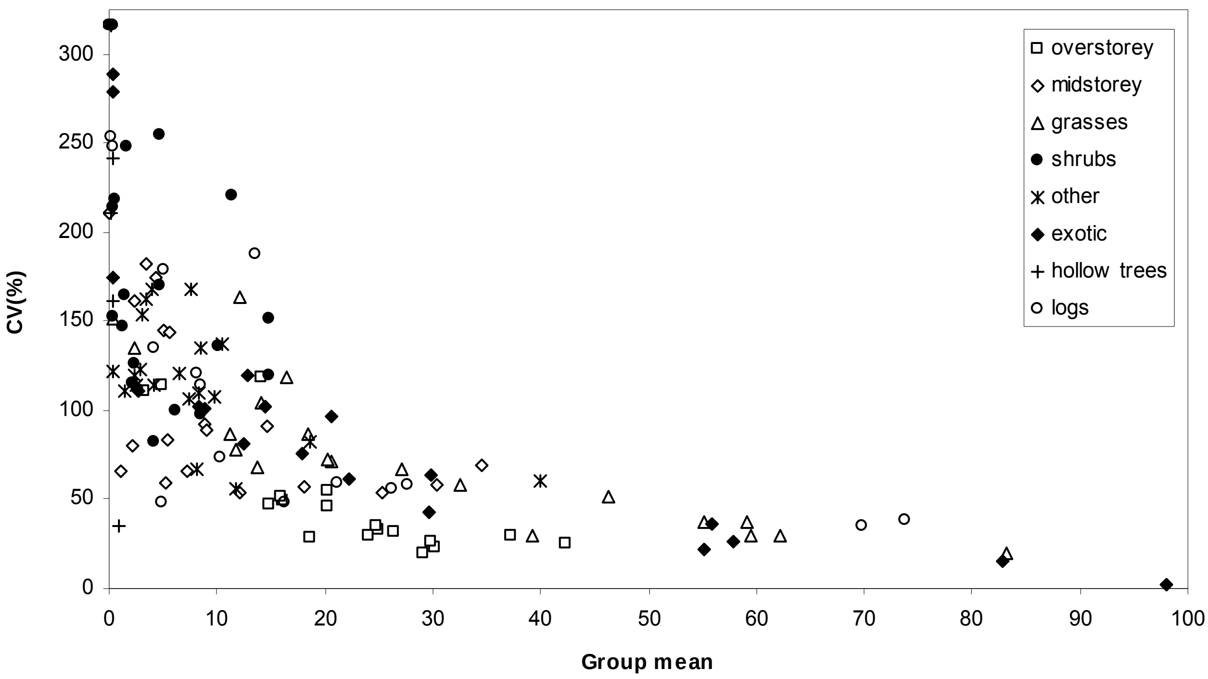
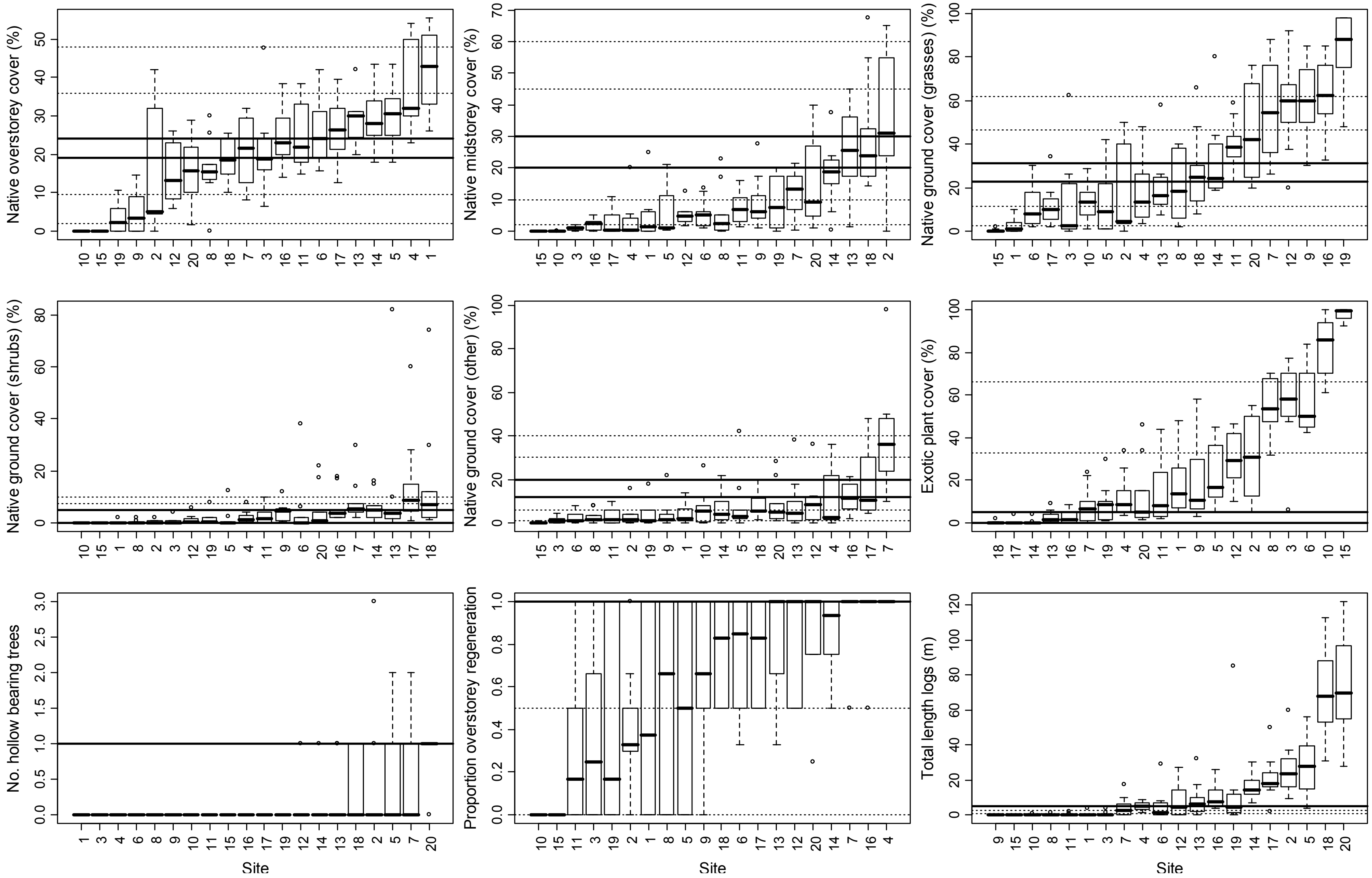


Figure 1. Coefficients of variation for vegetation attributes estimated by ten observers on twenty sites for BioMetric assessments.

The range of raw estimates for each attribute on each site almost always spread across multiple BioMetric scoring categories (Figure 2). Exceptions (in which all observations were within the same scoring category) were for most BioMetric attributes on the pasture sites (Sites 10 and 15); sites on which no observers recorded a *hollow bearing tree*; and sites for which mean values of *exotic plant cover*, *proportion of overstorey species regenerating* and *length of logs* were high (Figure 2). Outliers in estimates of *native ground cover (shrubs)* were primarily from one observer (Observer J). The BioMetric attributes for which raw estimates spread across the greatest number of scoring categories on average were: proportion of *overstorey species regenerating*, and the lowest was for *number of hollow bearing trees* (because there are only two categories).

Figure 2. Boxplots of observer's estimates of vegetation attributes included in the BioMetric vegetation condition metrics. Thick horizontal lines indicate the benchmark scoring category for the attribute, and thin horizontal lines indicate the other scoring categories. Sites are arranged in order of increasing mean value for the attribute. Boxplots show median, quartiles andoutliers.

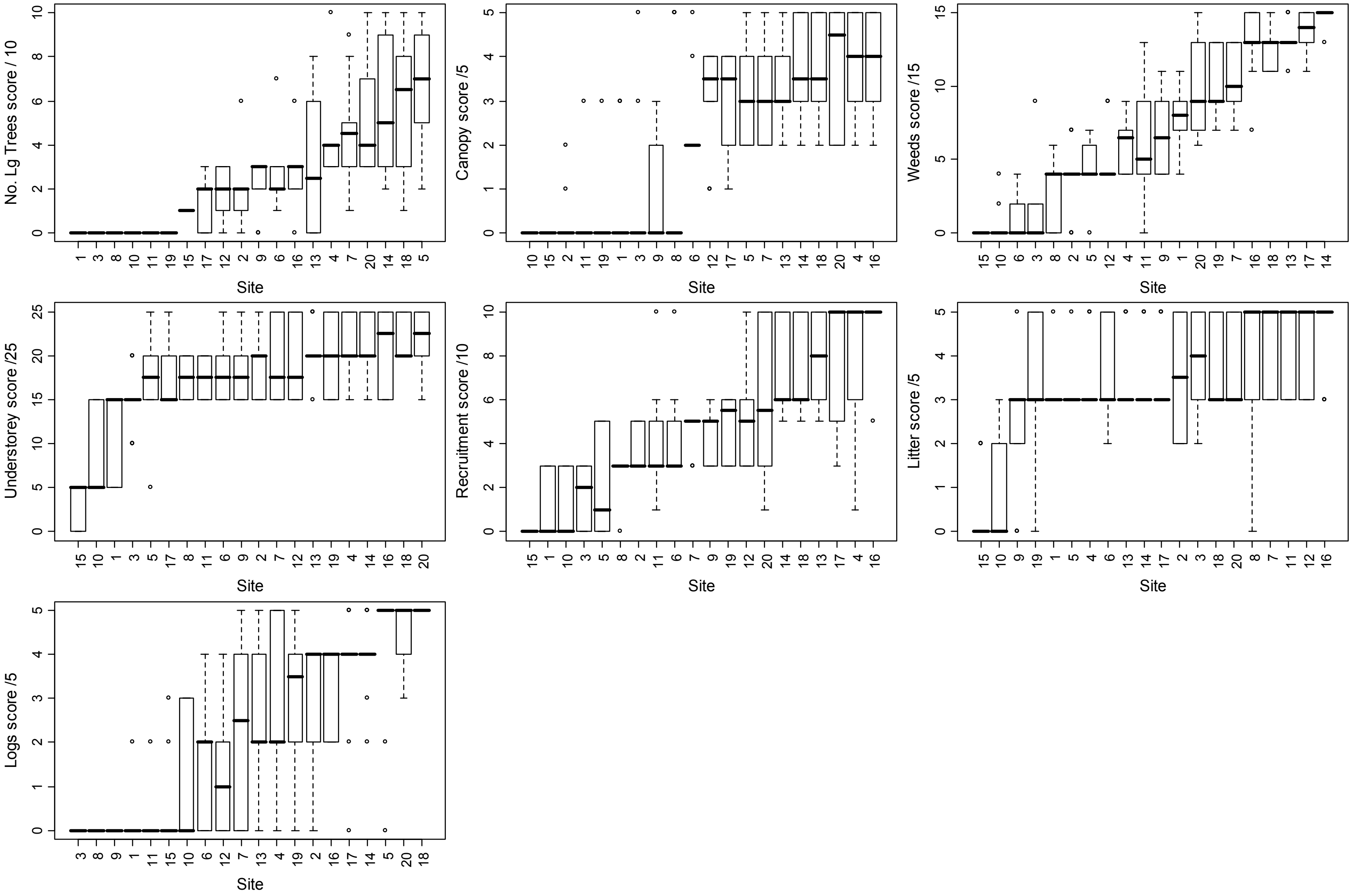


Observers' estimates also generally spanned multiple scoring categories for Habitat Hectares attributes, except for those pasture sites (10 and 15) and when there were no large trees, canopy (though not always), or logs (Figure 3). Observer estimates of Habitat Hectares' *understorey* score span ten points (two scoring categories) for all sites, except site 15. The other highly weighted attributes varied substantially too (but less than understorey), the range of scores for *number of large trees* increased as mean score increased (with range of 7-9 out of 10 points for the five sites with highest mean scores), *weeds* had greatest spread of scores for intermediate mean scores, and recruitment.

Though our study was not designed to distinguish between measurement and identification errors, the latter were apparent where some observers recorded *native overstorey cover* in the BioMetric assessment as zero (absent) when most observers recorded it as present (Figure 2).

BioMetric scoring structure may have exacerbated the expression of observer variation for some attributes. This was most apparent for *native ground cover (other)*, in which raw estimates below 10% cover that differed by only a few percentage points frequently spanned three scoring thresholds (Figure 2). For most attributes, including *native ground cover (grasses)* and *proportion of overstorey species regenerating* (Figure 2), raw estimates varied substantially and were scattered across multiple scoring thresholds. Point estimates were not recorded for Habitat Hectares.

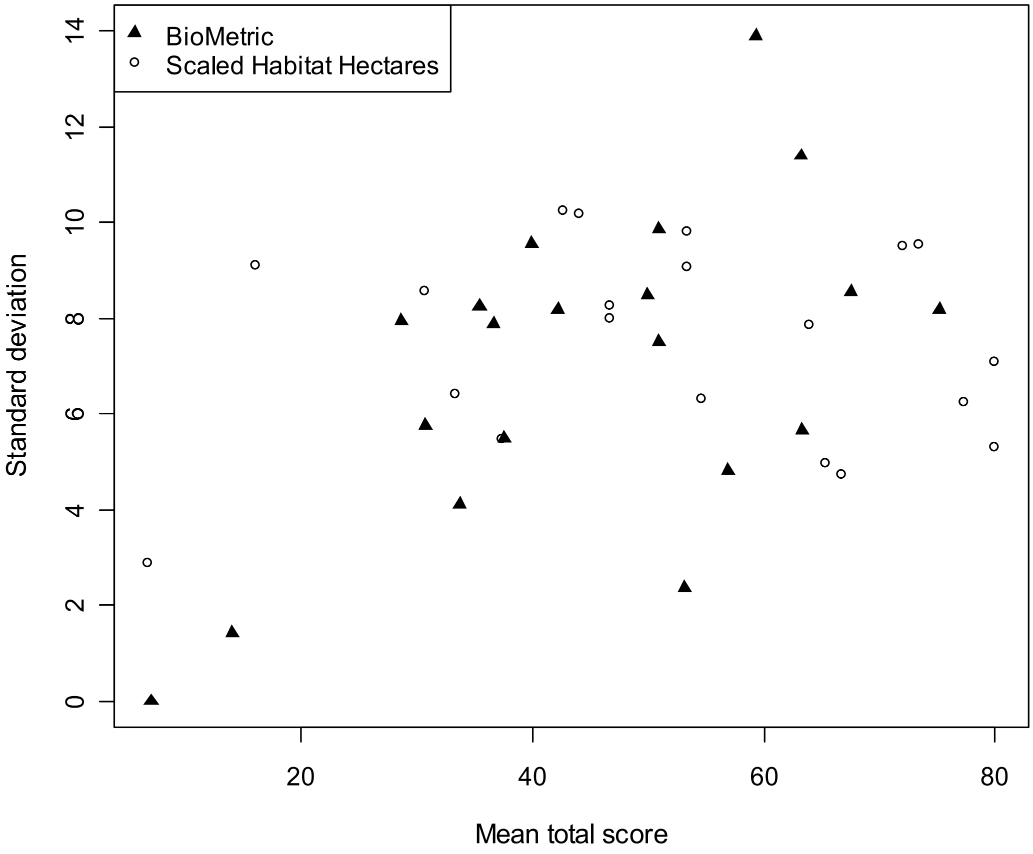
Figure 3. Boxplots of scores recorded by ten observers for Habitat Hectares assessments on 20 sites. Sites are arranged in order of increasing mean score for the attribute. Boxplots show median, quartiles and outliers.



Variation in total BioMetric and Habitat Hectares scores

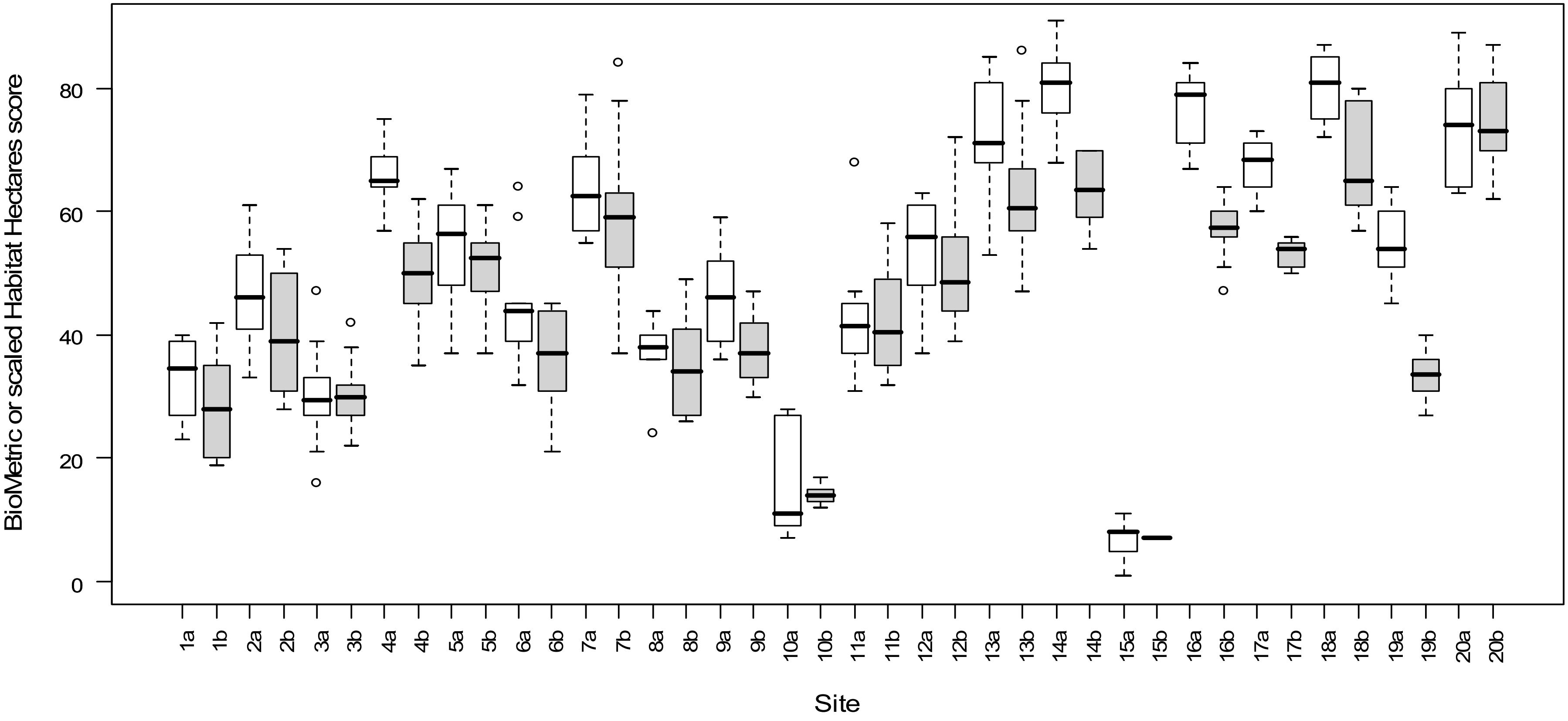
The average standard deviation around the mean across all sites was 7 for BioMetric and 5.5 for Habitat Hectares. For almost all sites with mean total BioMetric or Habitat Hectares scores greater than 30, the standard deviation was between approximately 4 and 10 (Figure 4). Two remnant sites had slightly higher standard deviations for BioMetric: Sites 7 and 13 had mean scores of about 60 and standard deviations of 14 and 11 respectively. The two pasture sites (Sites 10 and 15) had low mean scores and low standard deviations, with the exception of Site 10 for Habitat Hectares which had a higher standard deviation due largely to differences in *understorey* scores. The CV of total scores declined from approximately 30% to approximately 10% as mean scores increased from 30 to 80 for both metrics (data not shown). The average CV of total scores across observers was 15% for BioMetric and 18% for Habitat Hectares, and the maximum was 60% (Site 10 according to Habitat Hectares).

Figure 4. Standard deviations of total vegetation condition scores, as measured by ten observers using the BioMetric and Habitat Hectares methods.



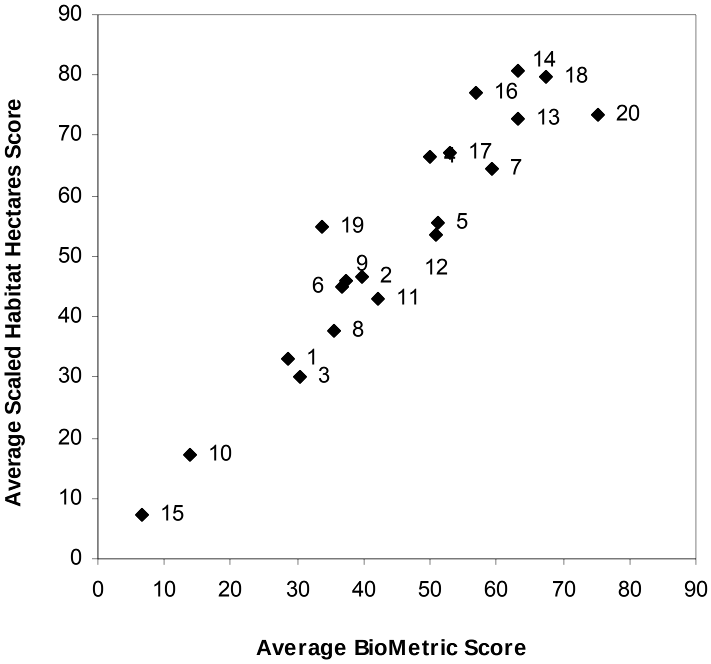
Variation in total scores for both BioMetric and Habitat Hectares were not explained by single factors. In some instances, variation arose from small errors between observers in heavily weighted attributes (e.g. *hollow bearing trees* in Sites 2, 7, 12 and 13 for BioMetric and *understorey* in Sites 5, 7, 10 and 20 for Habitat Hectares). In other cases, one or two observers estimated attributes very differently for multiple moderately weighted attributes (e.g. Site 11 for Habitat Hectares). In most cases, observer estimates varied above and below the group mean over all attributes consistent with random independent errors unexplained by observer experience, attribute or the details of field conditions.

Figure 5. Boxplots of total vegetation condition scores for each site, as measured by ten observers using the Habitat Hectares (open boxplots) and BioMetric (grey boxplots) methods. Boxplots show median, quartiles and outliers.



Based on the average total score across observers for each site, the BioMetric and Habitat Hectares protocols ranked sites similarly, with a Spearman correlation coefficient of 0.91. The lowest ranking sites in both methods lacked both a tree canopy and a shrub stratum, and had a ground stratum dominated by weeds (Table 3, Figure 6). The sites with low to moderate scores were either planted or lacked a tree canopy or native shrub stratum. The sites with higher scores in both methods were all remnant vegetation, and the highest ranking sites were structurally complex remnant vegetation with all three major strata and relatively few weeds (Table 3, Figure 6). No sites scored above an average of 80 points using either method.

Figure 6. Average BioMetric and scaled Habitat Hectares scores across observers for all sites.



Patterns amongst observers

The average rank correlation of all pairs of observers was calculated for each metric; as was the average correlation of each individual observer with all other observers. The average rank correlation across all pairs of observers was 0.82 (range 0.57 - 0.94) for BioMetric and 0.91 (range 0.82 - 0.97) for Habitat Hectares (Table 4). There was no apparent cause for the particularly low correlation between Observers G and H for BioMetric (0.57), and their Habitat Hectares correlation was high (0.92). The lowest rank correlation of an individual observer with other observers was 0.74 for Observer G, and this was only for BioMetric. All observers agreed on the lowest ranking site using BioMetric (Site 15), while two different sites were ranked lowest by observers using Habitat Hectares (Sites 10 and 15). There was less agreement amongst observers as to the ranks of moderate to high ranking sites using both protocols, with between 4 and 9 sites at each rank (data not shown).

Table 4. Spearman rank correlation coefficients of pairs of observers for total scores of BioMetric and *Habitat Hectares*.

	A		B		C		D		E		F		G		H		I		J	
B	0.93	*0.94*																		
C	0.80	*0.92*	0.88	*0.91*																
D	0.92	*0.89*	0.91	*0.93*	0.88	*0.84*														
E	0.86	*0.95*	0.87	*0.94*	0.74	*0.88*	0.85	*0.91*												
F	0.94	*0.88*	0.94	*0.96*	0.79	*0.83*	0.89	*0.93*	0.86	*0.89*										
G	0.73	*0.96*	0.78	*0.92*	0.82	*0.82*	0.75	*0.90*	0.69	*0.97*	0.74	*0.90*								
H	0.89	*0.93*	0.80	*0.92*	0.74	*0.90*	0.87	*0.91*	0.88	*0.94*	0.81	*0.88*	0.57#	*0.92*						
I	0.84	*0.93*	0.78	*0.90*	0.66	*0.92*	0.83	*0.85*	0.85	*0.93*	0.85	*0.84*	0.71	*0.89*	0.79	*0.93*				
J	0.85	*0.94*	0.81	*0.92*	0.82	*0.87*	0.86	*0.87*	0.86	*0.92*	0.80	*0.87*	0.84	*0.92*	0.86	*0.91*	0.81	*0.88*		
Observer Mean	**0.86**	***0.93***	**0.86**	***0.93***	**0.79**	***0.88***	**0.86**	***0.89***	**0.83**	***0.92***	**0.85**	***0.89***	**0.74**	***0.91***	**0.8**	***0.92***	**0.79**	***0.9***	**0.83**	***0.9***
Observer mean – Total mean	+ 0.04	+ *0.02*	+ 0.04	+ *0.02*	- 0.03	- *0.03*	+ 0.04	- *0.02*	+ 0.01	+ *0.01*	+ 0.03	- *0.02*	- 0.08	*0.00*	- 0.02	+ *0.01*	- 0.03	- *0.01*	+ 0.01	- *0.01*

#Not significantly different from zero, for two tailed probabilities at α = 0.01 (Zar 1972).

Only one observer's total site scores were, on average, consistently different from the group mean using both protocols: Observer B had slightly higher estimates (Figure 7). Each observer made the most extreme estimate (i.e. the absolute difference from the group mean, regardless of direction) on at least one site using BioMetric, except Observer A. Similarly, all observers made the most extreme estimate on at least one site using Habitat Hectares, except Observer E. The average absolute difference from the group mean was higher in Habitat Hectares than BioMetric.

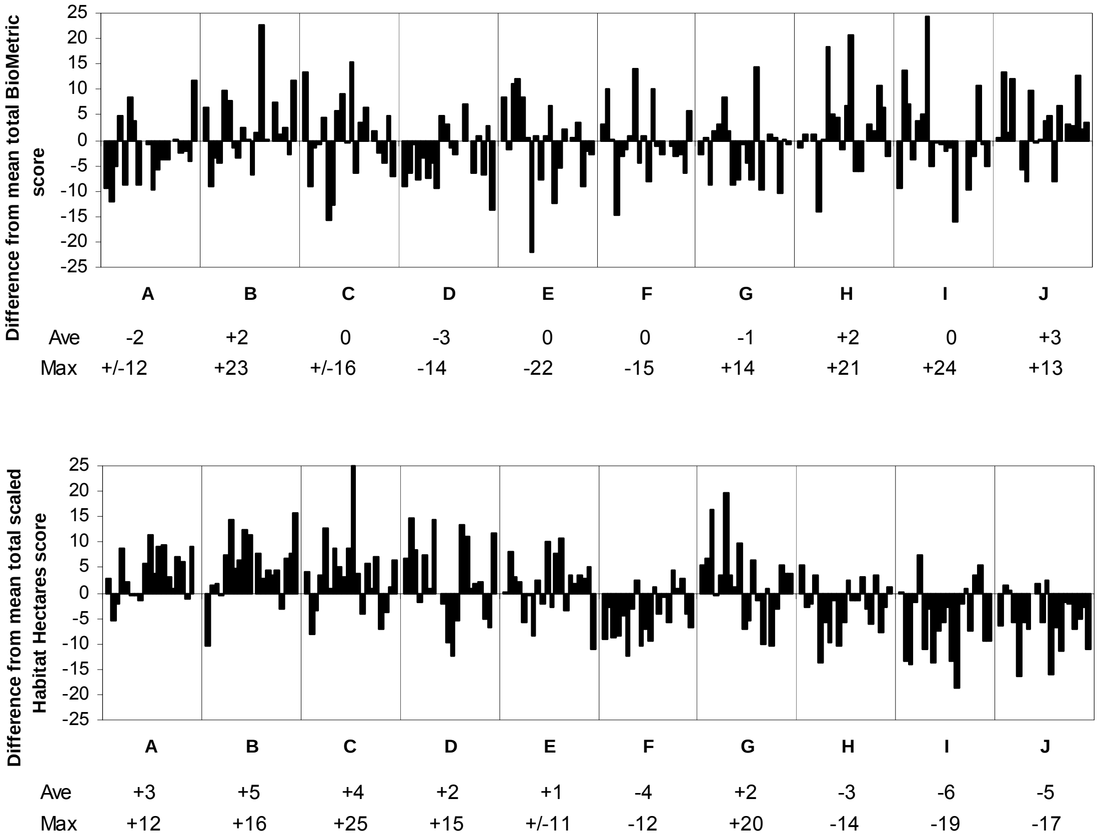


Figure 7. Difference of total a) BioMetric scores and b) scaled Habitat Hectares scores for Observers A-J from the group mean.

BioMetric and Habitat Hectares had quite different patterns of variation in estimates (Figure 7). Habitat Hectares had a larger proportion of variation between observers, with Observers A, B and C recording higher scores on average than the group mean and Observers H, I and J recording lower scores on average than the group mean (Figure 7). Using the BioMetric protocol, a greater proportion of variation occurred between sites within observers.

Discussion

Observer variation in vegetation condition scores

Observer variation in total scores was similar for BioMetric and Habitat Hectares, and does not appear to have been greatly affected by differences in the index structures or sampling methods. Variation in total scores was primarily caused by random observer variation in all vegetation attributes. The magnitude of variation detected (average CV of 15-18%) may have underestimated variation in real-world assessments as: observer variation in species richness was excluded; almost all observers were trained by the same person; all observers surveyed the exact same sites; and the order in which observers conducted assessments was non random.

There was general, but imperfect, agreement amongst observers about the rank order of sites. Most observers agreed as to the scores and ranks of very highly degraded (or poor condition) sites in which many attributes had zero values. At less degraded (moderate and high condition) sites, there was substantially more disagreement among observers as to site scores and ranks. Protocols for assessing vegetation condition therefore may not reliably distinguish between sites of moderate to high condition, though they should distinguish low condition sites with certainty.

Potential implications for decision making and biodiversity conservation

The results indicate that there may be considerable uncertainty as to which site is in best condition or whether any given site exceeds some threshold condition value for a management decision. In the context of offsetting decisions, underestimation of a development site and/or overestimation of an offset site may cause greater than expected loss of biodiversity. Suppose it is proposed to remove all vegetation from Site 14, and offset it with management actions on Site 4. Based on the average BioMetric scores of the two sites (50 and 63 respectively) and ignoring uncertainty in these estimates, management actions would be required to increase the value of Site 4 by 13 BioMetric points. However, the spread of observer scores about the mean values suggest that this decision could still lead to a loss of between 6 and 22 points (or alternatively a windfall of between 5 and 19 points) based on the highest and lowest estimated scores for these sites.

Dealing with observer variation: Increasing sample size

One possible method for dealing with observer variation in vegetation condition assessments for decision making would be to increase the sample size. Given observer errors, Block et al. (1987) calculated that sample sizes of greater than 75 may be required for precise point estimates for vegetation attributes. The results of this study indicate that ten observers may be insufficient to reliably estimate the biodiversity value of a site using two vegetation condition assessment protocols. Suppose decisions for offset assessment required vegetation condition to be estimated within 10% of the true value. Using Equation 2, the number of independent observers required would be up to 30 for both BioMetric and Habitat Hectares for most sites. Many hundreds of observers would be required for vegetation condition to be estimated within 1% of the true mean. Time and monetary constraints will therefore prevent the use of sample sizes sufficient to deal with observer variation.

$$\text{Number of observers required} = \frac{CV^2 \cdot t^2}{E^2} \qquad (2)$$

Where CV is the Coefficient of Variation, *t* is the standard value from the students t distribution and *E* is the standard error of the mean.

Dealing with observer variation: Improving precision

The scatter of extreme estimates across all observers suggested a tendency for any observer to make inaccurate estimates on occasion, and this was true regardless of the kind of attribute measured or the condition of the vegetation. These results emphasise the opportunity to improve operator performance by enhancing measurement protocols.

While empirical evidence of observer error has long been reported in relevant literature, conclusive evidence of causative factors and potential remedial actions is surprisingly scarce. Potential causes of variability within individual observers may include: complexity of the vegetation (an observer may be more or less accurate in particular vegetation structures or the vegetation may be heterogeneous); survey conditions such as weather and time of day (which may cause fatigue); and survey techniques (which may be subjective). Variability amongst different observers may be caused by different levels of training in the use of a particular survey technique or familiarity with a given vegetation structure.

There is some evidence that precision of observer's estimates may be improved, but not eradicated, through use of small sampling units (Leps and Handicova 1992; McCune and Lesica 1992; Klimes 2003; Archaux et al. 2007), or more objective survey techniques (Brakenhielm and Qinghong 1995; Zhou et al. 1998; Ringvall et al. 2005). Training may also improve precision, although evidence is limited and equivocal (Smith 1944; Sykes et al. 1983). The extent to which precision of observers can be improved through smaller sampling units, improved survey techniques (including minimising fatigue and reducing any linguistic uncertainty in operating manuals) and regular training and calibration against standards needs to be further researched.

Dealing with observer variation: Modifying scoring structures

Categorical scoring structures may have exacerbated the expression of observer error for some attributes. Moderate levels of observer variation across scoring categories in heavily weighted attributes can have very large impacts on total score. An example is the sensitivity of BioMetric to tree hollows. The variation in number of hollow bearing trees detected on any given site was low (maximum range of 0 to 3), yet total BioMetric score can be altered up to 22 percentage points for this attribute. These sensitivities to weighting suggest that, while observer error appears to be the main driver of uncertainty in vegetation condition scores, the heavy weighting of some attributes can exacerbate its effects.

Incorporating uncertainty into decision making

Assuming observer error is unlikely to be eradicated, and few independent observers are likely to be able to conduct assessments, estimates of vegetation condition may always be uncertain. Use of available insights on the magnitude and direction of uncertainty in field assessments can be explicitly considered in the decision making process, and will usually reduce the risk of decisions that lead to poor conservation outcomes. New methods for quantifying uncertainty and incorporating it into decision making processes are emerging (e.g. Ben-Haim 2001; Burgman et al. 2001). Most involve specifying a best estimate and describing the uncertainty around that estimate in the form of an interval, a fuzzy number or a probability distribution. Statistical probabilities should be used with caution in the absence of adequate knowledge of the nature of uncertainty. Valuable information about uncertainty in field assessments of vegetation condition may be obtained simply by requiring individual observers to provide bounded estimates for all parameters. Additional information may be obtained by requiring multiple observers to independently conduct assessments if there are Occupational Health and Safety requirements to work in teams.

Regardless of the method used to describe uncertainty around best estimates, uncertainty should be propagated through the calculation of vegetation condition scores. Taking the example of offsetting clearing of Site 14 with gains on Site 14, a more robust offset that minimised the risk of loss of biodiversity may require the biodiversity value of Site 4 to be increased by 35 points based on the highest estimated score for Site 14 and the lowest estimate for Site 4. More sophisticated methods could potentially be incorporated into software tools used to assist decision making, such as the BioMetric Decision Support Tool (DEC 2005b). Decisions may then be made that are more robust to uncertainty (unlikely to deliver unexpected bad outcomes), and less likely to cause loss of biodiversity than if the decisions were made on best estimates alone.

Conclusion

This study has shown appreciable levels of uncertainty in field assessments of vegetation condition, which may cause difficulty distinguishing between moderate and high value sites. Broadly similar levels of uncertainty were recorded for two protocols for assessing vegetation condition, due to imprecision in estimates of all vegetation attributes and potentially exacerbated by the protocols’ scoring structures. It is recommended that further research is conducted into methods for improving precision of observers’ estimates through regular training and calibration and reducing fatigue effects. It is also recommended that observers provide bounded estimates for all parameters, and uncertainty is formally incorporated into management decisions for biodiversity conservation.

References

Archaux F., Berges L., Chevalier R. (2007) Are plant censuses carried out on small quadrats more reliable than on larger ones? *Plant Ecology* 188, 179-190.

Ben-Haim Y. (2001) *Information-Gap Decision Theory: Decisions Under Severe Uncertainty*. Academic Press, London.

Block W.M., With K.A., Morrison M.L. (1987) On measuring bird habitat influence of observer variability and sample size. *The Condor* 89(2), 241-251.

Brakenhielm S. and Qinghong L. (1995) Comparison of field methods in vegetation monitoring. *Water, Air and Soil Pollution* 79, 75-87.

Burgman M.A., Breininger D.R., Duncan B.W., Ferson S. (2001) Setting reliability bounds on habitat suitability indices. *Ecological Applications* 11(1), 70-78.

Carlsson A.L.M., Bergfur J., Milberg P. (2005) Comparison of data from two vegetation monitoring methods in semi-natural grasslands. *Environmental Monitoring and Assessment* 100, 235-248.

DEC (2005a) BioBanking – *A biodiversity offsets and banking scheme. Conserving and restoring biodiversity in NSW. Working paper.* New South Wales Department of Environment and Conservation, Sydney. December 2005.

DEC (2005b) *BioMetric Decision Support Tool. Version 1.8 for the NSW Property Vegetation Plan Developer*. New South Wales Department of Environment and Conservation, Sydney. April 2005.

DEC (2006) *Reviewed interim vegetation condition benchmarks*. New South Wales Department of Environment and Conservation, Sydney. 8 August, 2006. Available from URL: http://www3.environment.nsw.gov.au/images/BioMetric_benchmark_data_reviewedAugust2006.xls. Accessed 25 March 2008.

DSE (2004) *Vegetation Quality Assessment Manual – Guidelines for applying the habitat hectares scoring method*. Version 1.3. Victorian Government Department of Sustainability and Environment, Melbourne.

DSE (2006) Bush Broker: *Native vegetation credit registration and trading. Information paper*. Victorian Government Department of Sustainability and Environment, Melbourne.

ESSS (2000) *Commonwealth listing advice on Cumberland Plain Woodlands*. Endangered Species Scientific Subcommittee, Australian Government Department of Environment, Water, Heritage and the Arts. Available from URL: http://www.environment.gov.au/biodiversity/threatened/communities/cumberland-plains.html. Accessed 25 March 2008.

Gibbons P., Freudenberger D. (2006) An overview of methods used to assess vegetation condition at the scale of the site. *Ecological Management and Restoration* 7(S1), S10-17.

Gibbons P., Ayers D., Seddon J., Doyle S., Briggs S. (2005) *BioMetric, A terrestrial biodiversity assessment tool for the NSW Property Vegetation Plan Developer, Operational manual*. Version 1.8. NSW Department of Environment and Conservation, c/- CSIRO Sustainable Ecosystems, Canberra.

Gorrod E., Keith D.A., Burgman M.A. (in review) Can free markets save biodiversity? An ecological perspective on biobanking. *Conservation Biology*

Grey A. N., Azuma D.L. (2005) Repeatability and implementation of a forest vegetation indicator. *Ecological Indicators* 5, 57-71.

Harper M.J., McCarthy M.A., van der Ree R., Fox, J.C. (2004) Overcoming bias in ground-based surveys of hollow bearing trees using double sampling. *Forest Ecology and Management* 190, 291-300.

Hellmann J.J. and Fowler G.W. (1990) Bias, precision and accuracy of four measures of species richness. *Ecological Applications* 9(3): 824-834.

Klimes L. (2003) Scale dependent variation in visual estimates of grassland plant cover. *Journal of Vegetation Science* 14, 815-821.

Leps J., Hadincova V. (1992) How reliable are our vegetation analyses? *Journal of Vegetation Science* 3, 119-124.

McCune B., Lesica P. (1992) The trade-off between species capture and quantitative accuracy in ecological inventory of lichens and bryophytes in forests in Montana. *The Bryologist* 95(3), 296-304.

NPWS (2004) *Endangered Ecological Community Information: Cumberland Plain Woodland*. New South Wales National Parks and Wildlife Service, Hurstville. February, 2004. Available from URL: http://www.environment.nsw.gov.au/resources/nature/EECinfoCumberlandPlainWoodland.pdf. Accessed 25 March 2008.

Oliver I., Ede A., Hawes W., Grieve A. (2005) The NSW Environmental Services Scheme: Results for the biodiversity benefits index, lessons learned and the way forward. *Ecological Management and Restoration* 6(3), 197 -205.

Oliver I., Jones H. and Schmoldt D.L. (2007) Expert panel assessment of attributes for natural variability benchmarks for biodiversity. *Austral Ecology* 32, 453-475.

Parkes D., Newell G., Cheal D. (2003) Assessing the quality of native vegetation: The 'habitat hectares' approach. *Ecological Management and Restoration* 4(supplement), S29-S38.

Regan H.M., Colyvan M., Burgman M.A. (2002) A taxonomy and treatment of uncertainty for ecology and ecological conservation. *Ecological Applications* 12(2), 618-628.

Ringvall A. Petersson H., Stahl G., Lamas T. (2005) Surveyor consistency in presence/absence sampling in monitoring vegetation in a boreal forest. *Forest Ecology and Mangement* 212, 109-117.

Smith A.D. (1944) A study of the reliability of range vegetation estimates. *Ecology* 25(4), 441-448.

Sykes J.M., Horrill A.D., Mountford M.D. (1983) Use of visual cover assessments as quantitative estimators of some British woodland taxa. *Journal of Ecology* 71, 437-450.

ten Kate K., Bishop J., Bayon R., (2004) Biodiversity offsets: Views, experience and the business case. IUCN, Gland, Switzerland and Insight Investment, London, United Kingdom. Available at URL: http://biodiversityeconomics.org/offsets. Accessed 25 March 2008.

Tozer M. (2003) The native vegetation of the Cumberland Plain, western Sydney: systematic classification and field identification of communities. *Cunninghamia* 8(1), 1-75.

USDA (2003) Conservation Reserve Program Sign-up 26: Environmental Benefits Index. United States Department of Agriculture, Farm Service Agency. May 2003. Available at URL: http://www.fsa.usda.gov/Internet/FSA_File/crpebi03.pdf. Accessed 25 March 2008.

van Hees W.W.S., Mead B.R. (2000) Ocular estimates of understorey vegetation structure in a closed *Picea glauca/Betula papyrifera* forest. *Journal of Vegetation Science* 11, 195-200.

Zhou Q., Robson M., Pilesjo P. (1998) On the ground estimation of vegetation cover in Australian Rangelands. *International Journal of Remote Sensing* 19(9), 1815-1820.