

Examining change over time in habitat attributes using Bayesian reinterpretation of categorical assessments

DAVID H. DUNCAN^{1,3} AND PETER A. VESK²

¹*Arthur Rylah Institute for Environmental Research, Department of Environment and Primary Industries, P.O. Box 137, Heidelberg, Victoria 3084 Australia*

²*Australian Research Council Centre of Excellence for Environmental Decisions, School of Botany, The University of Melbourne, Parkville, Victoria 3010 Australia*

Abstract. Prospects for evaluating effects of vegetation restoration have long been limited by availability of appropriately sensitive baseline data. Data that are typically collected to justify investment in restoration are rarely suitable for estimating subsequent change over time, but given how commonly such data are collected, can they contribute something to learning about ecological change over time? We compared vegetation and habitat data from a quantitative reassessment of 25 habitat restoration sites seven years after they were initially assessed using a semiquantitative, categorical scoring system. Our aim was to estimate the change at sites between the first, semiquantitative survey and a second, quantitative survey. We treated the initial values as effectively unknown and used Bayesian models to infer plausible values using three different informative prior distributions, variously comprising the initial site assessments and modeled values from a statewide data set. We successfully constructed models of change over time between the two surveys, and regardless of which prior model was implemented, our data analysis suggested that cover of exotic species was reduced, but canopy cover, the cover of organic litter, and the length of fallen logs were all increased after the seven-year period. A small increase in the mean number of large-diameter trees was likely due to initial measurement error. Site fertility and canopy cover were important covariates in explaining the magnitude of change in total log length. Sites with higher canopy cover decreased more in weed cover and increased more in litter cover. Our approach could be used to retrospectively analyze any ordinal data set where there is a scoring logic that can be interpreted quantitatively. Data sets where treatment contrasts and untreated controls exist will be particularly valuable for testing the utility of our approach. While this novel approach should prove a useful analytical complement to genuine longitudinal monitoring and space-for-time surveys, it is no substitute for initiation of learning about management effectiveness using data from purposefully designed and measured surveys.

Key words: *Bayesian statistics; categorical data; evaluating investment programs; incorporating prior data; inferring change; monitoring; payments for ecosystem services (PES) schemes; restoration; vegetation condition.*

INTRODUCTION

The inability of governments or other land managers to demonstrate return on investment in better management of the natural environment is attracting increasing scrutiny (Kleijn and Sutherland 2003, Bernhardt et al. 2005, Ferraro and Pattanayak 2006, Sutherland et al. 2006, Miteva et al. 2012). In Australia, successive audits of programs to tackle environmental degradation through native vegetation protection and restoration have highlighted the lack of evidence that these programs made an impact on their objectives (ANAO 2004, 2008). More broadly, the lack of monitoring and evaluation of anthropogenic impact on the environment, in restoration or degradation contexts, has been

highlighted for a number of decades (Underwood 1991, Legg and Nagy 2006, Lindenmayer and Likens 2010). Unless efficacy can be demonstrated, continuing public support for investment in native vegetation and habitat may be at risk (Duncan and Wintle 2008, Miteva et al. 2012).

Demonstrating the benefits of investment of public money on private land for improvement in the condition of native vegetation is a particularly difficult proposition. Uncertainty about ecological outcomes of standard actions such as fencing and livestock management is common to almost every scale, from single sites (Stoneham et al. 2003, Rumpff et al. 2011) to landscapes (Kyle and Duncan 2012). In this paper we focus on the site scale. Effective investment in vegetation change at sites would mean that the amount of change in vegetation variables targeted for management over time where investment occurs is demonstrably greater than non-investment sites. In the language of experimentation this

Manuscript received 1 October 2012; revised 31 January 2013; accepted 20 February 2013. Corresponding Editor: S. Findlay.

³ E-mail: david.duncan@depi.vic.gov.au

is best evaluated as a BACI design, Before (vs.) After; Control (vs.) Impact (Underwood 1991). These two analytical elements are fundamental to scientific inference; in this case because vegetation responds to both short-term disturbances that management may apply or alleviate, but also to external and legacy factors such as climate, and the intensity, frequency, and duration of past land use. Unfortunately, neither “before” nor “control” contrasts are typically measured in evaluating native vegetation investment programs. This is unfortunate because no change, or a slight decline in condition, might be good outcomes depending on what circumstances face noninvestment sites (Ferraro and Pattanayak 2006).

Many studies about vegetation change due to management have been drawn from space-for-time substitution studies, including many about grazing impacts (Yates et al. 2000, Floyd et al. 2003, Briggs et al. 2008, Read et al. 2011). Such studies are relatively strong on demonstrating an association between treatment and effect, but as we cannot know what the sites were like before the treatment, we learn little about rates of change per unit of time. Therefore, where an understanding of the potential response of specific sites is desirable, longitudinal data from sites through time will be particularly useful (Lindenmayer and Likens 2010). A major limitation in this regard has been the lack of appropriately sensitive baseline data (Taverna et al. 2005, Rumpff et al. 2009, Bottrill et al. 2011). Repeat sampling of sites using semiquantitative or subjective methods is unlikely to be informative, and data from new longitudinal studies accrues in real time; thus insights about change will be slow to materialize.

We address the problem of appropriate baseline data for vegetation composition and structure from 25 investment sites from the BushTender program in northern Victoria, Australia. When the program began in 2001, the vegetation at candidate sites was assessed using a semiquantitative method derived from coarse, categorical, subjective estimates. These types of data are ill suited to monitoring change over time because the coarse categorization leads to insensitivity, and for their subjectivity to inaccurate observation (Sykes et al. 1983, Gorrod and Keith 2009) and biases inherent in expert judgement (Burgman 2005, Speirs-Bridge et al. 2010). Further, no control sites were identified or assessed; thus our focus was limited to estimating change at investment sites, in particular vegetation or habitat measures at sites between the first, semiquantitative survey and the second, quantitative survey. We used Bayesian models to infer plausible initial values using informative prior distributions variously comprising the initial site assessments and modeled values from a statewide data set. Bayesian statistical approaches are increasingly employed in ecology to explicitly incorporate prior information in analyses of new data and coherently propagate uncertainty in models (McCarthy and Masters 2005, Hobbs and Hilborn 2006, Penman et al. 2009,

MacNeil and Graham 2010). However, we are not aware of previous studies that utilize Bayesian models to estimate plausible values from historical, categorical data. If the technique proved successful, there are many other coarsely quantitative or qualitative historical data sets that may yield important insights in ecology using similar approaches. Our specific aims were to demonstrate an approach to estimation of change by incorporating existing ordinal data into an analysis with new quantitative data, and to compare the influence of different prior distributions of the unknown initial scores on these inferences. With our preferred prior distributions we also explored whether selected site covariates plausibly helped explain the estimates of change.

METHODS

Study area, program and sites

BushTender was initiated during 2001 and 2002 by the Victorian Department of Sustainability and Environment (hereafter DSE, Stoneham et al. 2003). The trial, in which landholders were paid to apply or forsake management actions for habitat improvement, took place on remnant native vegetation on private properties in the foothills and plains of Northern Victoria. These landscapes were extensively cleared of native woody vegetation after discovery of alluvial gold in the 1850s, either for surface mining, timber to support mining infrastructure and settlements, or pastoral expansion influenced by goldrush optimism (Howitt 1855, Duncan et al. 2010). The more fertile parts were later used for grazing and cropping, while farming of some less fertile parts was abandoned. These less fertile landscapes were recolonized by secondary native forest and woodland communities after the early wave of clearing (Kyle and Duncan 2012). In recent decades, the area has seen considerable socio-economic transition away from primary production and toward rural residential use and hobby-farming (Barr et al. 2005). Under the BushTender trial, comprising 73 land holdings (131 “sites”), conservation management of nearly 3200 ha of native vegetation was secured for three years (Stoneham et al. 2003).

Data from initial site assessments (2001)

At the time when expressions of interest were being registered for the BushTender trial, site assessments were undertaken based on DSE’s vegetation quality methodology, “Habitat Hectares” (Parkes et al. 2003, 2004, DSE 2004, McCarthy et al. 2004). Assessments were based on units referred to as habitat zones (HZ); these were relatively homogenous areas within an Ecological Vegetation Class (EVC, Woodgate et al. 1994). Homogeneity in this context relates to relative degradation with reference to a benchmark state (Parkes et al. 2003); thus a vegetation type may be subdivided into habitat zones based on different levels of one or more condition components, for example, weediness. Once the HZ were

established, field workers made a roving semiquantitative assessment of the estimated number of mature trees >50 cm diameter at breast height, tree canopy cover, length of fallen logs, litter cover, weed cover, cover and diversity of understory life forms, and recruitment of native woody species (sensu DSE 2004). Unfortunately, no control or pseudo-control sites were identified. Thus, these data can only help us to learn what change occurred at sites where management was applied, not change relative to other treatments or to the continuation of the pre-existing management.

Site reassessments (2008)

From the suite of 131 sites with management contracts in the 2001 BushTender trial, we selected 25 habitat zones (HZ) on 18 properties for reassessment on the basis of landscape position and 2001 assessment scores. We sought zones on relatively fertile (plains) and infertile soil types (slopes), in relatively good and relatively poor starting condition. Starting condition was assigned on the basis of weed and understory scores; good sites in most cases had <5% weed cover, and poor sites >50% cover. Some zones on more fertile sites that had 25–50% weed cover were included as “good” if they had >50% of expected understory life-forms present and approaching benchmark cover and diversity values, whereas the same weed score on an infertile site was classed as “poor” if it had very low understory values (see DSE 2004:27). We attempted to minimize the differences in HZ area between good and poor zones. However, good zones tended to be larger in area than poor zones; the 75th percentile of the area of poor zones was about equal to the median value of the good zones. We sampled 10 different EVCs of similar structure, and sought those that were common throughout the study area. The HZ that we sampled were subject to agreements detailing explicit management commitments that landowners would retain all standing trees and fallen timber, and control pest herbivores and alien plant species where they had been identified as a concern. Livestock grazing was to be excluded from all but two HZ, in which extensive historic tree removal had replaced a woodland or forest community with extensive cover of derived pasture. The three-year BushTender management agreements had since expired, theoretically freeing participants to resume or assume land uses proscribed during the period of the agreement. Therefore we confirmed with landholders that the grazing management conditions of the trial had been continued for those HZ we intended to revisit.

At each site we first established a 100 × 40 m quadrat. The origin of the central transect was haphazardly located near the center of the HZ on an aerial image. When on site, we navigated to the location using differential GPS. If the quadrat could fit in any direction from the origin point, a random bearing was selected. If not, a random bearing was drawn from a restricted range or set of ranges. The quadrat was marked out

using GPS with waypoints marking each corner and the start and finish of the central transect. Within the quadrat, the following components were measured according to a method we had previously applied in similar vegetation types (Rumpff et al. 2009)

Large trees.—All trees >50 cm diameter at breast height (dbh) were measured. The initial assessment followed the manual’s instruction to count the number of “large” trees greater than a threshold diameter, specified as a benchmark. The applicable benchmarks for the EVCs sampled ranged from 60 to 80 cm dbh, but for our analysis we included trees down to 5 cm smaller than the benchmark value in our 2008 count to avoid excess zero estimates owing to small sample plots.

Tree canopy cover.—Ten digital images were taken at 10-m intervals along the 100 m long central transect, using a tripod-mounted Canon S40 digital camera (4.0 megapixels, 35 mm focal length) directed vertically with the aid of a spirit level. To derive a value for average projective canopy cover, the pixels in each captured image were allocated to either canopy or sky on the basis of color, using the proprietary software WINCAM (Regent Instruments, Quebec, Canada [2002]) and a set of training images with different levels of cloud and clear sky.

We used tree canopy cover estimates derived from aerial photography to satisfy ourselves that our 100 × 40 m subsamples were representative of the canopy cover of their parent HZ. We used image analysis software eCognition (Definiens eCognition Professional Version 4: Trimble, Westminster, Colorado, USA) to segment a GIS layer of the HZ into objects according to pixel values. These object polygons were superimposed over the red and infrared bands of the imagery to emphasize the contrast between canopy and noncanopy cover types. Polygons were then assigned manually to “canopy” or “noncanopy” classes. A proportional canopy cover was then calculated within the GIS for the HZ and the subsampled quadrat. Overall, the HZ and quadrat canopy cover estimates were closely related ($r^2 > 0.95$), thus ameliorating one major potential source of measurement error.

Logs.—The lengths of all logs (>10 cm diameter) within the quadrat were measured using vernier calipers and a tape measure and summed.

Understory structure.—The ground cover of organic litter and exotic species were recorded using point quadrats along five randomly selected 20-m subtransects running perpendicular to the main transect. At 20-cm intervals along the subtransect, each ground cover type and plant life form in contact with a vertical 1-m steel pin was recorded, providing 100 points per subtransect, or 500 points per HZ.

Modeling

Our aim in modeling was to estimate the change in particular vegetation or habitat measures at sites between the first, semiquantitative survey and second,

quantitative survey. We present the general model form and detail specifics of the models for the various attributes. We use Bayesian inference to estimate the model parameters. For each dependent variable (large trees, logs, tree canopy cover, weed cover, and litter cover), separately, we begin with an observation Y_{2i} from the second (quantitative) survey at site i . This observation was modeled as being drawn from some distribution appropriate to the three forms of data: proportional cover, length, and counts.

Here we describe the model for the change in a cover variable (canopy, weeds, and litter), before detailing specifics of models for the other variables. Cover values, represented as proportions, were first logit-transformed and then modeled as normally distributed. The logit(Y_{2i}) is drawn from a Normal distribution with grand mean μ , and standard deviation σ , which measures average departure of sites from μ , to be estimated from the data. The mean for site i , μ_i , is then modeled as a logical sum of the estimated mean of the i th site in the first survey, logit(Y_{1i}), and the change at the i th site, δ_i . We treat the measurement in the first survey Y_{1i} as an unknown to be estimated, with three alternative sets of informative priors (see *Priors*). We then write a linear model for the change at site i . Here, we model covariates as follows:

$$\delta_i = \alpha + \beta_1 \times X_1 + \beta_2 \times X_2$$

where α is the intercept and represents the average change at the hypothetical average site of low fertility. The β parameters represent regression coefficients, for the covariates X_1 representing the fertility (low or high) and X_2 projective cover of tree canopy (only for logs, litter, and weeds).

Variants.—The model to evaluate changes in the number of large trees was the simplest. Y_{2i} , being a count, was modeled as being drawn from a Poisson distribution with the single parameter μ_i for the mean and variance of site i . This mean for site i , μ_i , was then modeled as a logical sum of the estimated mean of the i th site in the first survey, logit(Y_{1i}), and the change at the i th site, δ_i .

The model to evaluate changes in the length of logs at a site begins with a transformation of observations by natural logarithms. The $\ln(Y_{2i})$ are drawn from a Normal distribution with mean μ_i , and standard deviation σ , to be estimated from the data. And μ_i is the logical sum of the estimated mean of the i th site in the first survey, $\ln(Y_{1i})$, and the change at the i th site, δ_i .

Priors

Bayesian inference requires explicit statement of prior probability distributions. We explored three different informative prior distribution forms for the first surveys for each dependent variable, variously emphasizing the original assessments and plausible distributions modeled from a statewide data set. The process of generating prior distributions that acknowledged the original

assessment involved a number of steps, as follows. The ordinal categorical scores from 2001 were first converted to their implied numeric ranges following the assessment manual (DSE 2004). The translation required three elements. The description of benchmark (reference) conditions for each EVC was sourced from DSE (2010). Second, the detailed scoring instructions for each assessed vegetation attribute, which explicitly require comparison against the relevant EVC benchmark, were obtained from the manual. Third, we assumed that the observer was correct in their categorical score of the HZ. This assumption was simply a practical one. It is clear that visual estimates of vegetation attributes over large scales are prone to several sources of bias and epistemic and linguistic uncertainty (Regan et al. 2002, Gorrod and Keith 2009). Thus, it is probable that different observers might have scored a given site differently (agreement), and it is also unclear how successful any one observer might be in estimating the true underlying value (accuracy). Nonetheless, this was a simple, moderately conservative assumption under the circumstances.

In the first informative prior distribution form, hereafter referred to as the Score prior, point values for each variable in each HZ were drawn from a uniform distribution bounded by the implied range described in the previous step, i.e., $Y_{1i} \sim \text{Uniform}(r_{1i}, r_{2i})$, where r_1 and r_2 define the range implied by the Assessors' score. For example, if the 2001 categorical weed score suggested a range in weed cover of 5–25%, values between 5 and 25 were drawn into the model from a uniform distribution. This reflects a lack of knowledge beyond the particular assessment. In the second informative prior distribution form, we ignored the implied range from the assessment and instead utilized distribution parameters estimated from neural network ensemble models linking categorical and quantitative scoring of the components in similar vegetation assemblages, based on data from southern Victoria (M. White, P. Griffioen, and G. Newell, *unpublished data*, and Appendix). These parameter estimates from the statewide models, according to the EVC of the HZ, we then treat as fixed parameters for our analyses of change; $Y_{1i} \sim \text{Beta}(a_i, b_i)$ for the cover values; $Y_{1i} \sim \text{Poisson}(\lambda_i)$ for the number of large trees; $Y_{1i} \sim \text{lognormal}(m_i, p_i)$ for the length of log. This we refer to as the Statewide prior. For the third informative prior distribution form we combined both previous approaches in that the implied range of the assessment was used to truncate the distribution generated from the neural network, resulting in a nonuniform prior, e.g., $Y_{1i} \sim \text{Beta}(a_i, b_i)$ truncated by the limits (r_{1i}, r_{2i}) for the cover values. We refer to this as the Combined prior.

All regression intercepts and coefficients were supplied vague priors (means of 0 and standard deviations of 1000). The parameter σ was modeled as uniformly distributed between 0 and 100. All analyses were implemented using OpenBUGS Version 3.1.2 (Spiegel-

halter et al. 2005). Inference was based on three Monte Carlo Markov Chains for each model. Convergence, when all three chains sampled similar ranges, was confirmed by eye to occur within the first 10 000 iterations; these were discarded as “burn-in” (e.g., McCarthy 2007) and our inference about the posteriors was based on the following 100 000 iterations.

Model evaluation

We found no model evaluation precedent that suited our study, in which estimated change was the focus of our inference. Popular summary evaluation measures, such as deviance explained or Deviance Information Criterion (Spiegelhalter et al. 2002), could not be reliably estimated for all our models, and anyhow lack specificity to inference about the estimated change. Therefore, our first and simplest approach to model evaluation was to look for credible results. For example, no change in the number of large trees was expected, so we were more likely to trust models that conformed to that assumption. Secondly, we tested the sensitivity of the modeled change to the choice of prior distribution form by comparing the estimated mean change, α , for an intercept-only model. We assumed that an intercept-only model would have relatively low sensitivity to the choice of prior. Because we had not yet modeled variation in the change, the uncertainty around the mean change should be reasonably large, relative to the difference due to the prior. Then covariates were added. As more covariates are added to a model, because some of the variation in the response was explained, the uncertainty around the mean change was reduced. Also, the models became “tuned,” and so more sensitive to the choice of prior. From our exploration of our model evaluation options two distinct modes of sensitivity were apparent: first, the “best model” of a candidate set may differ depending upon the choice of prior; and second, the estimate of the mean change for any given explanatory model may vary between the different prior forms, to a greater degree than for an intercept-only model. Because our data set lacked a convincing reference point for the “true” change that we were estimating, we could not objectively appraise model performance and sensitivity in either mode. Therefore we ultimately retreated to a conservative approach to model evaluation, as described previously, and to implementation of explanatory models. Using just one prior distribution form, the Combined prior, we explored candidate models for each response variable using a suite of covariates, yet present just one set of models, using one or two of the most obvious covariates. Our aim was not to find the “best model” for each response, but rather to illustrate the potential to explain variation in the estimated change.

RESULTS

The Score, Statewide, and Combined prior probability distributions had very different shapes (Fig. 1a), and

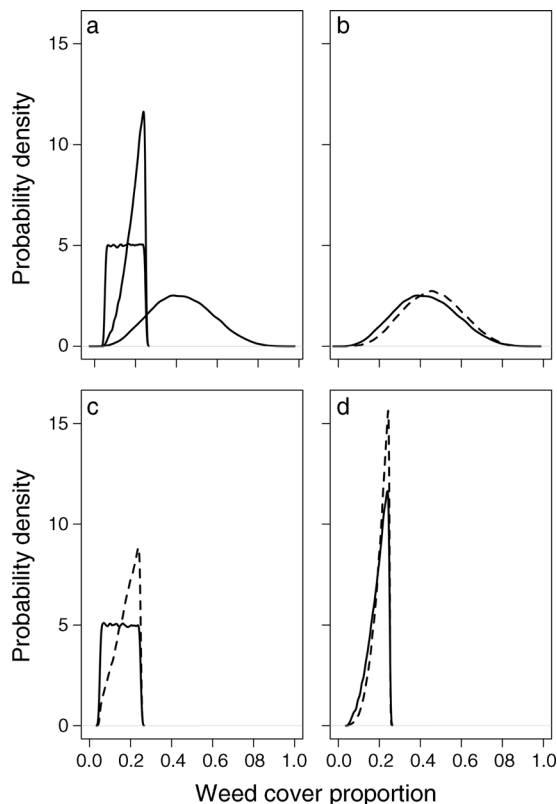


FIG. 1. Example comparison of the prior and posterior probability distribution forms for one initial weed cover assessment. In panel (a) the three prior distribution forms are compared. In subsequent panels the prior (solid) and posterior (dashed) distributions are compared for (b) Statewide (parameters estimated from statewide models for the vegetation type), (c) Score (parameters drawn from a uniform distribution truncated by the categorical score definition), and (d) Combination (the Statewide form truncated by the category definition) prior model forms.

resulted in differently shaped posteriors for the first surveys (Fig. 1b–d). However, in most cases we found that the sign and the magnitude of α in the intercept models were not greatly influenced by the choice of prior (Fig. 2). The mean estimate for weed cover was consistently lower in 2008, and with similar uncertainty regardless of the prior, although the credible intervals for each overlapped with zero. Canopy cover was higher for two of the three prior model forms, and the length of logs and litter cover were higher in 2008. For the count of large trees the credible intervals encompassed no change for the Score prior, whereas the Statewide and Combined priors suggested that the number of large trees was two trees greater on average.

In the passages that follow, to explore results for change in individual vegetation and habitat attributes for each HZ, we use only the Combined prior (prior 3), which assumed that: (1) the assessor’s score was correct, and (2) that within the truncated distribution given by the score, the value was likely to have been similar to

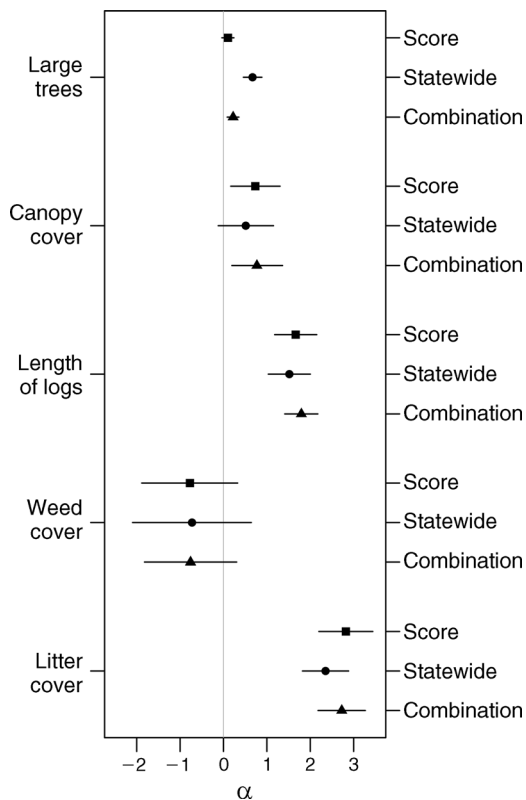


FIG. 2. The effect of choice of prior distribution form on α , the parameter for the change (mean and 95% credible intervals). The habitat attributes (response variables) are shown along the left side the graph.

sites of the same size range, in the same EVC, elsewhere in Victoria. Given that the comparisons of prior distribution forms indicated no clearly superior form, we chose the Combined prior because we preferred the logic that unknown values drew from both the specific local and general statewide data.

Number of large trees

We expected no change in the number of large trees. But the mean effect using the Combined prior (3) was a slight increase, with a negative β estimate for the effect of fertility indicating that the increases were associated with low-nutrient sites (Fig. 3). The data for individual HZ (Fig. 4b) suggested that the overall positive effect was driven by one strongly positive outlier, where the 2001 survey data indicated 0 large trees, whereas our 2008 census suggested 28 large trees/ha. In that instance there must have been an error either in the initial zone mapping or zone assessment, as changes of such magnitude over this time scale are implausible. Increases were only suggested in 3 out of 24 cases across all HZ, and for 6 out of 11 higher fertility HZ (Fig. 4a), the estimated number of large trees was lower in 2008.

Tree canopy cover estimates

Canopy cover estimates for 2008, based on canopy photographs, were higher on average than those from 2001, which were based on visual estimates (Fig. 3). The posterior parameter estimate for the effect of fertility indicated no difference between high- and low-nutrient sites (Fig. 3). The individual HZ across high- and low-fertility HZ (Fig. 4c, d) show that in many cases the 2008 data overlapped with the initial score range from 2001, but not with the posterior probability distribution for 2001. Decreased canopy cover was not suggested for any HZ, although in fact we observed considerable mortality of adult *Eucalyptus macrorhynca* F. Muell. ex Benth. in the field, particularly in low-fertility sites, and might have expected to see a decline were the same assessment techniques applied in both surveys.

Fallen timber (logs >10 cm diameter)

The amount of fallen timber on the ground in 2008 was estimated to be higher on average than in 2001, with increases of similar magnitude for high- and low-fertility HZ (Fig. 3). The individual HZ data at first glance seemed to suggest greater increases on lower-fertility sites (Fig. 4f) than higher-fertility sites (Fig. 4e), but on closer inspection the lower-fertility HZ are simply more heterogeneous, with fallen timber in low-fertility HZ increasing little, and higher-fertility HZ increasing strongly.

Cover of weed species

Weed cover was typically lower in 2008 compared with 2001 (Fig. 3). The effect was greater on lower-fertility sites and where canopy cover was higher (Fig. 3). Decreases appeared to have occurred mostly among

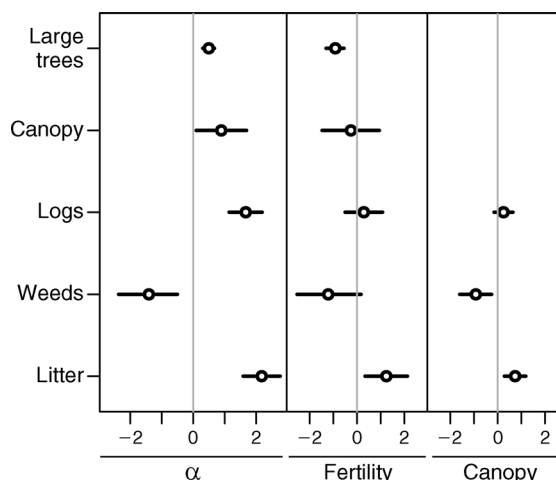


FIG. 3. Estimate of model terms (and 95% credible intervals) from a standardized model structure where change in large trees and canopy was a function of fertility, and change in logs, weeds, and litter was a function of canopy cover and fertility.

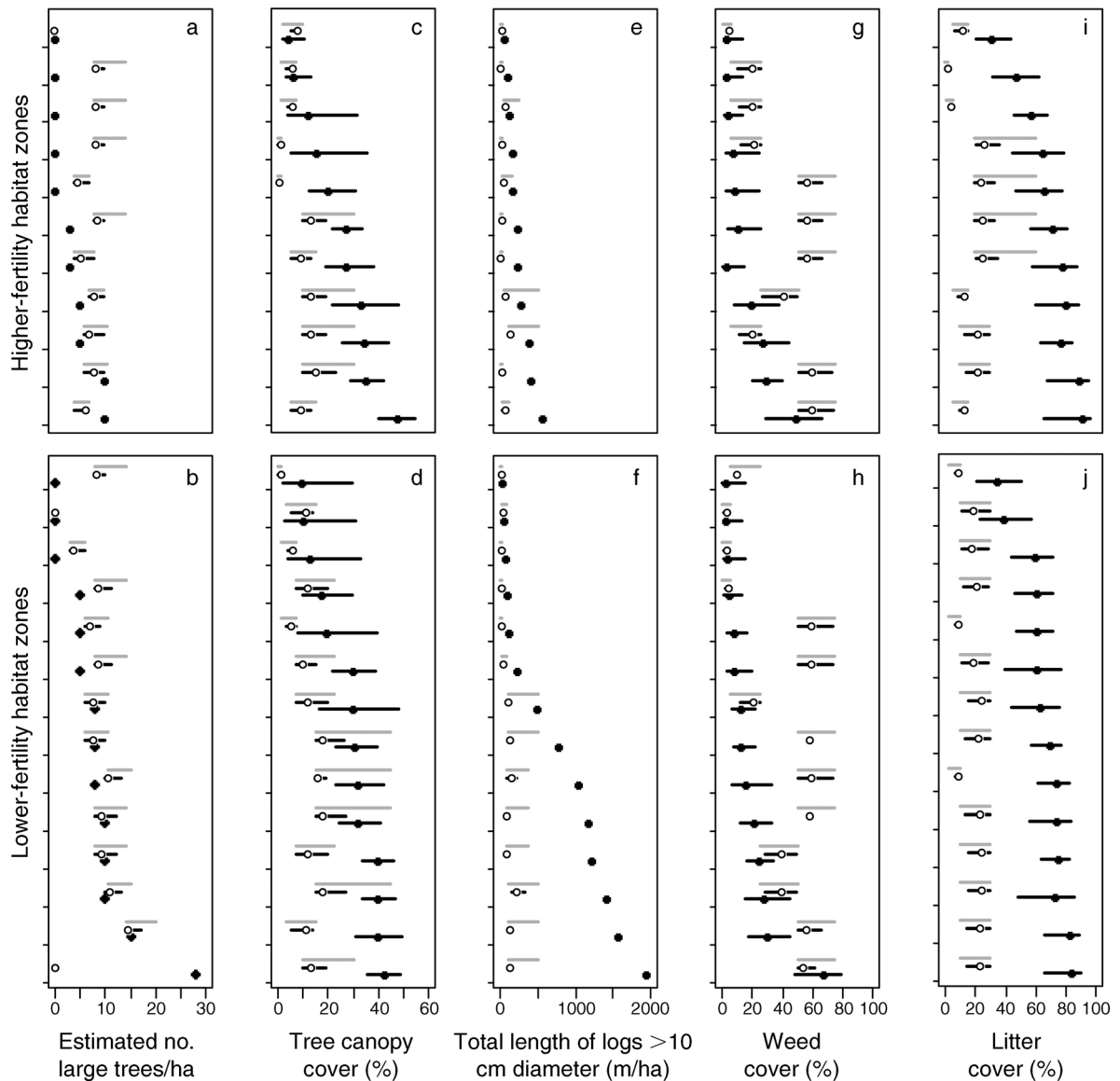


FIG. 4. Estimated change in (a, b) large trees, (c, d) tree canopy cover, (e, f) logs, (g, h) weed cover, and (i, j) organic litter for sites of higher (upper row) and lower (lower row) fertility. Modeled census and point-quadrat field data from 2008 (solid symbols and 95% credible intervals) are shown against two versions of the 2001 data: the implied numerical range from original categorical assessments (gray lines); and the modeled posterior distribution of the mean 2001 values based on the Combined prior (open symbols and 95% credible intervals). In each panel, habitat zones are ordered according to the estimated mean value for 2008; hence their location differs from panel to panel.

the sites that were scored as having relatively higher (>50%) weed cover in the 2001 assessment (Fig. 4d).

estimate and the posterior estimate for 2001 did not overlap (Fig. 4e).

Organic litter cover (%)

DISCUSSION

Organic litter cover estimates were substantially higher for most sites in 2008 than in 2001, and the change was greater at higher-fertility sites, and where higher canopy cover values were recorded (Fig. 3). The average increase was by >50 percentage cover units, or 4.5 times the initial cover estimate, to an average of 70% ± 17% cover (mean ± SD), and for most cases the 2008

We have demonstrated quantitative estimation of change in vegetation and habitat variables through time by modeling plausible values from historical semiquantitative data. This enabled the evaluation of ecological outcomes associated with the pilot BushTender conservation management program sites. We now discuss some of the implied change in variables, the caveats that

should be borne in mind with this analysis, and ways that it might be improved upon in the future.

From this set of 25 woodland and dry-forest sites, a consistent pattern emerged of reduced weed cover and accumulation of the structural habitat elements of litter and fallen timber. The observed changes likely represent a mixture of management regime and climatic conditions. For the decrease in weed cover, the anecdotal evidence from landholders (Victorian Department of Sustainability and Environment, *unpublished landholder reports*) suggests the decrease was influenced by the combination of herbicide application and prolonged dry conditions experienced in southeastern Australia between 2001 and 2010 (Leblanc et al. 2009, Verdon-Kidd and Kiem 2009), which probably suppressed annual weeds' germination and growth. Most landholders reported that they carried out the weed removal treatments required under their contracts, though some said they did not need to implement control in later years because weed cover was already reduced by the drought conditions. It would be interesting to know if the apparent reduction in weed cover was sustained once the drought was broken in 2010 and followed by two years of high spring and summer rainfall.

Equally, the drought period probably contributed to accumulation of litter and logs at these HZ through water stress on tree leaves and limbs. As noted earlier, mortality of adult trees was commonly observed in dry forest sites in particular and could have contributed to significant litter and timber fall. Rumpff et al. (2009), using the same field sampling technique as employed in our 2008 surveys, found that 13 unburned Grassy Dry Forest sites on public land also increased litter cover from an average of 50% to almost 80% over five years between 2003 and 2008. By way of contrast, however, in that same study, no increase in logs was observed for unburned sites, whereas our data suggest a large increase.

The magnitude of change in each component can only be cautiously estimated from these models, as the initial site estimates used different methods at different scales from those employed at the follow-up assessment. In the next section, we discuss specific examples where we feel that the results hinted at structure in observer error. Nonetheless, if one assumes that observer and estimation errors were either random, or systematic but relatively consistent across all sites, then the findings from the models are useful nonetheless. Furthermore, without conducting an analysis such as this, such idiosyncrasies of assessment and scoring as may be suggested by our analysis may not have come to light.

Legacy of sampling and estimation error

It is likely that individual HZ changes in large tree count, increases in canopy cover, and the accumulation of fallen timber, were strongly influenced by measurement error in 2001. Our a priori assumption was that minimal if any change in large trees would have

occurred between assessments. Growth in mature trees should be too slow to result in a detectable increase over seven years, and tree felling resulting in a decrease was unlikely given that the sites were under conservation management agreements. Certainly, no physical sign of tree felling or natural fall was evident. Therefore, the differences likely reflected under- or overestimation at the first assessment due to (1) poorly calibrated visual estimates of girth leading to failure to notice "large trees," or (2) difficulty in estimating the number of large trees in a large HZ. Also, particularly for large trees, it is possible that both assessments were correct and that differences were related to the mismatch of sampling scale between roving, whole-of-HZ assessments in 2001, and fixed plots in 2008. While it is plausible that canopy cover could have changed, it is also clear that among canopy estimation techniques, ocular estimates and those from digital photographs are relatively prone to bias and variance (Korhonen et al. 2006).

For the amount of logs on the ground, while an increase was likely, the implied magnitude of change from these data, with increases of up to 2 km/ha of material on the ground, are scarcely credible. Such increases could only have occurred with considerable felling or import, which did not happen at these sites. Log data for 2008 was a census from the 4000-m² plot for which low measurement error was expected; therefore the most likely source of error is underestimation of the amount of fallen timber in the 2001 assessments. This may have been further compounded by the applicable benchmark log lengths, which were 150–200 m/ha for the sampled EVCs. Theoretically, in 2001 assessors scored HZ for logs according to the percentage of benchmark length observed, but we speculate that perhaps if most sites were easily going to achieve the maximum categorical score ($\geq 50\%$ of benchmark length), assessors might subconsciously, intuitively standardize sites or "grade against the curve," rather than being guided strictly by the rules of the scoring method.

By contrast, the original weed cover scores, which are not assessed by comparison with a benchmark, seemed more plausible. Benchmarks serve an important purpose in conservation planning, allowing comparison of relative naturalness across diverse vegetation communities (Hunter 1996, Landres et al. 1999, Oliver et al. 2002, Gibbons et al. 2010). However, if benchmark values induce assessors to make systematic estimation errors, then they may need to be separated from the field assessment procedure and reserved for use as an interpretative filter. In Victoria, the State Government now requires consultant assessors to major vegetation condition survey projects to use a Personal Digital Assistant (PDA) or other mobile device, which makes the benchmark comparison automatically after the assessor provides their visual estimate of quantity. Although there is reason to suspect that relieving assessors of the additional calculation may reduce

observer error, this has not yet been demonstrated to our knowledge.

A novel application of Bayesian analysis to interpret prior data?

We believe that this work constitutes a novel application of Bayesian statistics and modeling in applied plant ecology. We have been able to evaluate support for change in vegetation structure associated with management agreements by comparing quantitative, posttreatment data with unknown, pretreatment values drawn from informative prior distributions. The key ingredient was a traceable scoring logic for the categorical scores for each component in the metric from which implied quantitative intervals can be recovered. These were further informed using models based on the strength of other data from a comparable cohort of sites elsewhere in the state. Those two elements were combined with detailed new assessments to analyze likely change over time in the structure of vegetation at these investment sites. In fact, the modeled output that we drew on to refine the prior data (the unpublished model [by M. White, P. Griffioen, and G. Newell] linking point estimates of vegetation condition components to resulting categories, summarized in the Appendix) exploited a similar logic. They used neural network ensemble models to learn the relationship between point estimates and categorical scores, given ancillary environmental and terrain data, so that they could predict point values for vegetation components for sites where only categorical data had been collected. Those analyses had different methods, scale, and purpose to the current study, but both highlight how coarse categorical data can be reanalyzed to create useful map and model representations of native vegetation condition and change.

The analyses described here tested one specific assumption about the historical data set. We gave the assessors the benefit of the doubt for correctly assigning habitat zones to the ordinal score that encompassed the true value at that time. However, further sophistications in specifying the informative prior distribution forms may be possible, including allowing for some probability of misassignment of the categorical score (Spiegelhalter and Stovin 1983, Milne and Walter 1998, Royle and Link 2005). For example, if most values between 0 and 25% cover occur close to the upper bound, it seems probable that spillage into the adjacent score would make the interpretation of historical data both more conservative and realistic.

Detecting change at investment sites vs. demonstrating the impact of investments

The preceding discussion reflects on this attempt to tease out the likely ecological change at these sites. The strength of this work is in using available evidence to extract a change narrative from unknown pretreatment values among considerable measurement error. Howev-

er, these data could not have shed light on the effectiveness of the investments in a causal sense, even if all of the limitations described above were successfully addressed. In order to quantify the benefit of the investment, the change at treated sites must be considered in the light of what are other sites are doing (controls). It needs to be recognized that so-called “counterfactual evidence” from controls is as fundamental to justifying management and investment regimes (Ferraro and Pattanayak 2006) as it is to scientific experimentation.

Identification of appropriate controls for public investment in improved management of native vegetation on private land is not a trivial task, nor is it in any conservation policy context (Miteva et al. 2012). Such controls would ideally be similar to native vegetation remnants on private land where no investment was occurring. One readily identifiable source of noninvestment “controls” could be to work with sites and landholders who were “losers” from bids or funding requests. These are known sites for which some assessment data may have been collected through the funding application process. If those data suggest substantial overlap in site condition, then such landholders and sites may be worth pursuing. However, part of the logic of extension and incentive programs is that, of itself, the interaction that occurs between government agents and landholders during the bid development process may encourage practice change (Windle et al. 2009). If this is true, then “failed” bidders may not constitute true controls if they begin to manage their sites differently despite not being funded. Clearly, identifying appropriate controls or pseudo controls will be challenging, but it is fundamental to learning about management effectiveness.

Conservation typically gets a meager budget allocation, so how difficult might it be to win support for spending more time and money on monitoring and reporting the effectiveness of conservation activities? This tension about allocation of resources may become moot. In the light of increasing adoption of offset policies (e.g., ten Kate et al. 2004) and market-based investment in environmental services (Engel et al. 2008, Gómez-Baggethun et al. 2010, Yang et al. 2010), rapid improvement of standards of monitoring and evidence may be unavoidable. Even though the nature of the majority of activities undertaken at sites have barely changed over the past few decades (e.g., fencing; seeding and planting of native species), the formality of the new transactions such as legally binding and outcome-based contracts (e.g., Gibbons et al. 2011), their monetary value, and even their philosophical basis, carry implicit demands for higher standards of evidence. Whereas the typical conservation agreement a decade ago was an informal negotiation between an interested landholder and an extension agent, the increasingly formal market-based or market-like approaches (e.g., Yang et al. 2010)

will inevitably drive demand for more rigorous standards of evidence.

CONCLUSIONS

We have demonstrated a Bayesian modeling approach that enabled categorical assessments to be utilized in estimating quantitative change over time in vegetation and habitat variables. The model structure could be employed wherever a trail of logic links categorical assessments to a continuous quantitative scale. Ultimately, however, these findings argue for better data to support monitoring of program effectiveness, including a coherent plan for counterfactual evidence, either by establishing pseudo controls or complementary space-for-time approaches. These elements should ideally be driven by, designed, and organized in light of clearly articulated objectives within an adaptive management framework.

ACKNOWLEDGMENTS

This work was made possible by 19 landholders, field assessors from 2001 and particularly 2008, including head office and data support. Elizabeth Farmer conducted image segmentation in eCognition on our behalf, and Mick McCarthy contributed to the early development of the Bayesian models. Earlier versions of this document were improved by the comments of Libby Rumpff, Matt White, Michael Scroggie, and two anonymous referees.

This work was funded by the Australian Government under the Commonwealth Environmental Research Facility Landscape Logic, and the Victorian State Government Department of Sustainability and Environment. P. Vesik was supported by the Australian Research Council Centre of Excellence for Environmental Decisions and the Australian Government National Research Hub on Environmental Decisions. The views in this paper are those of the authors and do not necessarily represent those of the State of Victoria or the Australian Government.

LITERATURE CITED

- ANAO. 2004. The administration of the National Action Plan For Salinity And Water Quality. Australian National Audit Office, Canberra, Australia.
- ANAO. 2008. Regional delivery model for the Natural Heritage Trust and the National Action Plan For Salinity And Water Quality. Australian National Audit Office, Canberra, Australia.
- Barr, N., R. Wilkinson, and K. Karunaratne. 2005. Understanding rural Victoria. Victorian Government Department of Primary Industries, Melbourne, Australia.
- Bernhardt, E. S., M. A. Palmer, J. D. Allan, G. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, and J. Follstad-Shah. 2005. Synthesizing U.S. river restoration efforts. *Science* 308:636.
- Bottrill, M., M. Hockings, and H. P. Possingham. 2011. In pursuit of knowledge: addressing barriers to effective conservation evaluation. *Ecology and Society* 16:14.
- Briggs, S. V., N. M. Taws, J. A. Seddon, and B. Vanzella. 2008. Condition of fenced and unfenced remnant vegetation in inland catchments in south-eastern Australia. *Australian Journal of Botany* 56:590–599.
- Burgman, M. A. 2005. Risk and decisions for conservation and environmental management. Cambridge University Press, Cambridge, UK.
- DSE. 2004. Vegetation quality assessment manual: guidelines for applying the habitat hectares scoring method. Version 1.3. Victorian Government Department of Sustainability and Environment, East Melbourne, Australia.
- DSE. 2010. Ecological Vegetation Class (EVC) benchmarks for each bioregion. Victorian Government Department of Sustainability and Environment, East Melbourne, Australia. <http://www.dse.vic.gov.au/conservation-and-environment/native-vegetation-groups-for-victoria/ecological-vegetation-class-evc-benchmarks-by-bioregion>
- Duncan, D. H., G. Kyle, and D. Race. 2010. Combining facilitated dialogue and spatial data analysis to compile landscape history. *Environmental Conservation* 37:432–441.
- Duncan, D. H., and B. A. Wintle. 2008. Towards adaptive management of native vegetation in regional landscapes. Pages 159–182 in C. Pettit, W. Cartwright, I. Bishop, K. Lowell, D. Pullar, and D. Duncan, editors. *Landscape analysis and visualisation*. Springer Verlag, Berlin, Germany.
- Engel, S., S. Pagiola, and S. Wunder. 2008. Designing payments for environmental services in theory and practice: an overview of the issues. *Ecological Economics* 65:663–674.
- Ferraro, P. J., and S. K. Pattanayak. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology* 4:e105.
- Floyd, M. L., T. L. Fleischner, D. Hanna, and P. Whitefield. 2003. Effects of historic livestock grazing on vegetation at Chaco Culture National Historic Park, New Mexico. *Conservation Biology* 17:1703–1711.
- Gibbons, J. M., E. Nicholson, E. J. Milner-Gulland, and J. P. G. Jones. 2011. Should payments for biodiversity conservation be based on action or results? *Journal of Applied Ecology* 48:1218–1226.
- Gibbons, P., S. V. Briggs, D. Y. Murphy, D. B. Lindenmayer, C. McElhinny, and M. Brookhouse. 2010. Benchmark stem densities for forests and woodlands in south-eastern Australia under conditions of relatively little modification by humans since European settlement. *Forest Ecology and Management* 260:2125–2133.
- Gómez-Baggethun, E., R. De Groot, P. L. Lomas, and C. Montes. 2010. The history of ecosystem services in economic theory and practice: from early notions to markets and payment schemes. *Ecological Economics* 69:1209–1218.
- Gorrod, E. J., and D. A. Keith. 2009. Observer variation in field assessments of vegetation condition: implications for biodiversity conservation. *Ecological Management and Restoration* 10:31–40.
- Hobbs, N. T., and R. Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. *Ecological Applications* 16:5–19.
- Howitt, W. 1855. Land, labour and gold; or two years in Victoria with visits to Sydney and Van Diemen's Land. Longman, Brown, Green and Longman, London, UK.
- Hunter, M. 1996. Editorial: benchmarks for managing ecosystems: are human activities natural? *Conservation Biology* 10:695–697.
- Kleijn, D., and W. J. Sutherland. 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity? *Journal of Applied Ecology* 40:947–969.
- Korhonen, L., K. T. Korhonen, M. Rautiainen, and P. Stenberg. 2006. Estimation of forest canopy cover: a comparison of field measurement techniques. *Silva Fennica* 40:577–588.
- Kyle, G., and D. H. Duncan. 2012. Arresting the rate of land clearing: change in woody native vegetation cover in a changing agricultural landscape. *Landscape and Urban Planning* 106:165–173.
- Landres, P. B., P. Morgan, and F. J. Swanson. 1999. Overview of the use of natural variability concepts in managing ecological systems. *Ecological Applications* 9:1179–1188.
- Leblanc, M. J., P. Tregoning, G. Ramillien, S. O. Tweed, and A. Fakes. 2009. Basin-scale, integrated observations of the

- early 21st century multiyear drought in southeast Australia. *Water Resources Research* 45:1–10.
- Legg, C. J., and L. Nagy. 2006. Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management* 78:194–199.
- Lindenmayer, D. B., and G. E. Likens. 2010. The science and application of ecological monitoring. *Biological Conservation* 143:1317–1328.
- MacNeil, M. A., and N. A. J. Graham. 2010. Enabling regional management in a changing climate through Bayesian meta-analysis of a large-scale disturbance. *Global Ecology and Biogeography* 19:412–421.
- McCarthy, M. A. 2007. *Bayesian methods for ecology*. Cambridge University Press, Cambridge, Massachusetts, USA.
- McCarthy, M. A., and P. Masters. 2005. Profiting from prior information in Bayesian analyses of ecological data. *Ecology* 42:1012–1019.
- McCarthy, M. A., et al. 2004. The habitat hectares approach to vegetation assessment: an evaluation and suggestions for improvement. *Ecological Management and Restoration* 5:24–27.
- Milne, M., and G. H. Walter. 1998. Host species and plant part specificity of the polyphagous onion thrips, *Thrips tabaci* Lindeman (Thysanoptera: Thripidae), in an Australian cotton-growing area. *Australian Journal of Entomology* 37:115–119.
- Miteva, D. A., S. K. Pattanayak, and P. J. Ferraro. 2012. Evaluation of biodiversity policy instruments: what works and what doesn't? *Oxford Review of Economic Policy* 28:69–92.
- Oliver, I., P. L. Smith, I. Lunt, and D. Parkes. 2002. Pre-1750 vegetation, naturalness and vegetation condition: What are the implications for biodiversity conservation? *Ecological Management and Restoration* 3:176–178.
- Parkes, D., G. Newell, and D. Cheal. 2003. Assessing the quality of native vegetation: the “habitat hectares” approach. *Ecological Management and Restoration* 4(Supplement):S29–S38.
- Parkes, D., G. Newell, and D. Cheal. 2004. The development and raison d'être of “habitat hectares”: a response to McCarthy et al. (2004). *Ecological Management and Restoration* 5:28–29.
- Penman, T. D., D. L. Binns, and R. P. Kavanagh. 2009. Quantifying successional changes in response to forest disturbances. *Applied Vegetation Science* 11:261–268.
- Read, C. F., D. H. Duncan, P. A. Vesk, and J. Elith. 2011. Surprisingly fast recovery of biological soil crusts following livestock removal in southern Australia. *Journal of Vegetation Science* 22:905–916.
- Regan, H. M., M. Colyvan, and M. A. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12:618–628.
- Royle, J. A., and W. A. Link. 2005. A general class of multinomial mixture models for anuran calling survey data. *Ecology* 86:2505–2512.
- Rumpff, L., D. H. Duncan, and A. Tolsma. 2009. Detecting short-term change in the structure and composition of dry grassy forest. Australian Government Department of Environment, Water, Heritage and the Arts, Hobart, Australia.
- Rumpff, L., D. H. Duncan, P. A. Vesk, D. A. Keith, and B. A. Wintle. 2011. State-and-transition modelling for adaptive management of native woodlands. *Biological Conservation* 144:1224–1236.
- Speirs-Bridge, A., F. Fidler, M. McBride, L. Flander, G. Cumming, and M. Burgman. 2010. Reducing overconfidence in the interval judgements of experts. *Risk Analysis* 30:512–523.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 64:583–639.
- Spiegelhalter, D. J., and P. G. I. Stovin. 1983. An analysis of repeated biopsies following cardiac transplantation. *Statistics in Medicine* 2:33–40.
- Spiegelhalter, D. J., A. Thomas, N. Best, and D. Lunn. 2005. *WinBUGS User Manual Version 2.10*. MRC Biostatistics Unit, Cambridge, UK.
- Stoneham, G., V. Chaudhri, A. Ha, and L. Strappazon. 2003. Auctions for conservation contracts: an empirical examination of Victoria's BushTender trial. *Australian Journal of Agricultural and Resource Economics* 47:477–500.
- Sutherland, W. J., et al. 2006. The identification of 100 ecological questions of high policy relevance in the UK. *Journal of Applied Ecology* 43:617–627.
- Sykes, J., A. Horrill, and M. Mountford. 1983. Use of visual cover estimates as quantitative estimators of some British woodland taxa. *Journal of Ecology* 71:437–450.
- Taverna, K., R. K. Peet, and L. C. Phillips. 2005. Long-term change in ground-layer vegetation of deciduous forests of the North Carolina Piedmont, USA. *Journal of Ecology* 93:202–213.
- ten Kate, K., J. Bishop, and R. Bayon. 2004. Biodiversity offsets: views, experience, and the business case. IUCN, Gland, Switzerland.
- Underwood, A. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Marine and Freshwater Research* 42:569–587.
- Verdon-Kidd, D. C., and A. S. Kiem. 2009. Nature and causes of protracted droughts in southeast Australia: comparison between the Federation, WWII, and Big Dry droughts. *Geophysical Research Letters* 36:L22707.
- Windle, J., J. Rolfe, J. McCosker, and A. Lingard. 2009. A conservation auction for landscape linkage in the southern Desert Uplands, Queensland. *Rangeland Journal* 31:127–135.
- Woodgate, P. W., W. D. Peel, K. T. Ritman, J. E. Coram, A. Brady, A. J. Rule, and J. C. G. Banks. 1994. A study of the old growth forests of East Gippsland. Department of Conservation and Natural Resources, Melbourne, Australia.
- Yang, W., B. A. Bryan, D. H. MacDonald, J. R. Ward, G. Wells, N. D. Crossman, and J. D. Connor. 2010. A conservation industry for sustaining natural capital and ecosystem services in agricultural landscapes. *Ecological Economics* 69:680–689.
- Yates, C. J., D. A. Norton, and R. J. Hobbs. 2000. Grazing effects on plant cover, soil and microclimate in fragmented woodlands in south-western Australia: implications for restoration. *Austral Ecology* 25:36–47.

SUPPLEMENTAL MATERIAL

Appendix

A description of unpublished neural network models of native vegetation and habitat variables ([Ecological Archives A023-066-A1](#)).



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Duncan, DH; Vesk, PA

Title:

Examining change over time in habitat attributes using Bayesian reinterpretation of categorical assessments

Date:

2013-09-01

Citation:

Duncan, D. H. & Vesk, P. A. (2013). Examining change over time in habitat attributes using Bayesian reinterpretation of categorical assessments. *ECOLOGICAL APPLICATIONS*, 23 (6), pp.1277-1287. <https://doi.org/10.1890/12-1670.1>.

Persistent Link:

<http://hdl.handle.net/11343/216870>

File Description:

Published version