# Creating an environmental commodity: the impact of alternative metric structures on relative investment values in the Environmental Stewardship Program in Australia

**Authors:**

Stuart M. Whitten, CSIRO Ecosystem Sciences, GPO Box 1700, Canberra ACT 2601.
Ph. +61 62464359. E: stuart.whitten@csiro.au

Art Langston. CSIRO Ecosystem Sciences, art.langston@csiro.au

**Working paper May 2013 submitted for consideration for BioEcon 15. Please do not cite or quote without permission of authors. All errors remain the responsibility of the authors.**

**Abstract**

Created environmental markets such as conservation tenders and biodiversity offsets employ a metric as a way of comparing the relative environmental value offered by alternative options in the market. Largely driven by ecological theory, these metrics have received little attention by economists yet are the critical linchpin that defines the nature of the values that are traded in markets. In this paper we describe some of the features required by a suitable metric in an environmental market. We then examine what the implications of deviation from these features are for the efficiency of one environmental market – the Australian Government's Environmental Stewardship Program – via a case study of the different metric formats employed as the program has evolved from 2008 to 2012. Our preliminary results suggest that there is little divergence in the efficiency or other measures of performance for different benefit estimates within a single metric functional form. In contrast, different functional forms, even when closely related, deliver substantive differences in investment efficiency across two metrics examined. These results are sufficiently encouraging tom motivate deeper analysis across a larger combined data set and three different metric forms.

**Keywords:** Conservation tenders; metrics; fungibility; benefit indices

**JEL codes:** Q57, Q58, D47

# 1. Introduction

Over the last thirty years there has been increasing focus on market-based mechanisms for delivering environmental benefits. These mechanisms focus on using prices to signal and incentivize agents, usually landholders, to manage resources under their management in order to protect or enhance environmental value they generate. Mechanisms employed span conservation tenders or auctions, offset mechanisms, payment for ecosystem service (PES) schemes and others. Most of these market-based mechanisms employ a metric to translate the quantity of environmental value provided into a commodity unit or score. Examples include the environmental benefits indexes applied within the Conservation Reserve Program in the United States (Ribaudo *et al.* 2001; Cattaneo *et al.* 2006; Soil & Water Conservation Society and Environmental Defense Fund 2008), differing iterations of the conservation value scores applied in the Australian Government Environmental Stewardship Scheme (Whitten *et al.* 2010), and variations on the habitat hectares scoring system applied in BushTender and BushBroker in Victoria, Australia (Parkes *et al.* 2003).

Metrics in environmental markets are the critical connection between the desired outcome and the strength of market signal provided (Boyd and Banzhaf 2007). The metric quantifies the environmental value as a unit that is intended to allow the comparison between different bundles of environmental services being offered in the market. As such they provide the critical role of fungibility in environmental markets without which they could not function. Without consistently quantified unit of investment value potential purchasers are unable to compare the benefits that they will receive from alternative offers of supply. Similarly, in a mature market, an agreed unit of exchange allows potential service providers to assess alternative investment decisions in delivering environmental benefits.

Conservation tenders have seen a wide range of metrics applied in different applications. Yet to date there has been relatively little research into the impact of different metric structures and forms on either the services provided or the market outcome. Recent examination of benefit estimates by for example Maron *et al.* (2013), risk (Moilanen *et al.* 2009), or from an economic perspective (Zabel and Roe 2009), does not shed much light on the way in which metrics should be constructed from an economic structural point of view, nor does much of this draw directly on early conceptual work in the economics domain by for example (Babcock *et al.* 1997; Wu *et al.* 2001). Similarly, index construction theory (Ebert and Welsch 2004) and practice (Ribaudo, Hoag *et al.* 2001; Boyd and Banzhaf 2007; Hajkowicz *et al.* 2009) tends toward an analysis of the attributes included rather than the influence of structure. There has been some research into the field implementation of metrics (Gorrod and Keith 2009; Cook *et al.* 2010) which is likely to influence attribute selection but not construction. And finally, there has been some evaluative oriented research undertaken into the efficiency gains from applying a metric compared to gains from cost pressure (Feather and Hellerstein 1997; Stoneham *et al.* 2003; Claassen *et al.* 2008; Connor *et al.* 2008).

To date there has been almost no research comparing the application of alternative forms of metric in a conservation tender using a consistent data set across the forms (the only attempt we are aware of is Rolfe and Windle 2011 though there are almost certainly others). Do different metrics really result in significantly different bundles of conservation benefits being purchased? There are two aspects to this question. First, does the type of benefit estimate (total value or net change to value otherwise known as additionality) impact on the bundle of contracts purchased? Second, does variation in the structure of the metric (biodiversity condition, expected value, multi-criteria index etc) impact on the bundle of contracts purchased? These are important questions as construction of metrics is potentially costly, yet likely to be critical to market performance. Hence, research into the appropriateness and effectiveness of metrics is essential to analysis of market performance as well as evaluating the payoff from investments in metric design and implementation.

The objectives of this paper are threefold. First, to identify the fundamentals of metric design applied in conservation tenders. Second, to identify the performance criteria that an optimal metric should meet. These are primarily derived from a review of relevant theory and practice. Third, to illustrate the potential implications of divergence from these performance criteria for investment decisions. To this end we compare the bundles of successful bids using similar rules across a modified form of the three metric structures employed in the Australian Government Environmental Stewardship Program (ESP) for the period 2008 to 2012.

The paper is structured as follows. An overview of the role, theory and experience in metrics implemented in conservation tenders is set out in Section 2. A case study of the metrics applied in one such tender, the Environmental Stewardship Program, is set out in Section 3, along with a set of research questions and an analysis method. The results are described in Section 4. The paper finishes with a discussion and brief summary of conclusions in Section 5.

# 2. Metrics in practice – experience in environmental tenders

## 2.1 The metric in a conservation tender

The use of metrics to guide the allocation of public funds via a conservation tender requires a specific approach that is different to that informing policy and program goals. In conservation tenders metrics are intended to describe the values that would result from alternate investments; they are the linkage between individual on-ground projects and the government investment program. A metric used in a competitive tender should capture the principal biophysical changes likely to result from management change at each site and in combination, and reflect the values that society places on the outcomes from these changes. The ideal metric would be:

- Scientifically robust: simple in structure while capturing the principal biophysical interrelationships and adequately quantifying how the system is likely to respond to different investments;
- Repeatable: generating the same estimate of investment value each time it is used for the same site and management actions;
- Transferable and replicable: capable of being implemented across the desired application domain and generating the equivalent scores for equivalent sites, in terms of the attributes captured in the metric;
- Economically and socially defensible: quantifying options in a way that reflects society's values;
- Ratio-scale comparability: the ratio difference in scores is a sufficiently accurate estimate of the difference in relative investment values to support decisions about relative funding levels (i.e. twice the score is twice as valuable from an investment point of view) (Ebert and Welsch 2004).

These principles translate into a metric that will enable quantitative comparisons across the range of potential options within the desired scale and scope of investment, as set by the specific policy and program goals of the investment. The ratio-scale comparability requirement is in contrast to an ordinal index in which we can only be confident that the best ranked bid is better than the next, not by how much. The latter can be used in a tender process, though less effectively, and requires a rank order analysis to evaluate bids. In all cases it is desirable for the metric to give the same score when applied to a singly defined vegetation patch, or to that patch divided into multiple units.

The composition of a metric should differ depending on the application. Biodiversity metrics for example usually comprise a measure of the conservation value (conservation value score), the scale

of the asset (usually area) and the impacts of different management actions, security or protection and potentially other factors. Attributes included within the metric will usually be a proxy or surrogate for an underlying desired value. For example, any measure of quantity and quality will necessarily be limited to an incomplete subset of the factors valued. Where possible the surrogate measure should be as closely correlated to the desired value as expense and practicality allow. Risk of poor estimates can also be reduced via the inclusion of different surrogates across different scales (Gibbons *et al.* 2009) or because they are more easily measured at different scales, but caution should be exercised to avoid double counting.

Any approaches will necessarily have some weaknesses as a result of knowledge and data gaps, scientific uncertainties, time until outcomes are delivered relative to entry to the program, practical implementation considerations, and constrained time and resources for metric development. Developing suitable metrics remains a complex task and is likely to remain so for some time. A close examination of any metric is likely to reveal a number of points where theoretical requirements are not met or practical imperatives resulted in a less than perfect approach. Nevertheless, paraphrasing McCarthy *et al.*'s (2004) rather critical review of one metric, careful judgements in developing metrics are likely to lead to investment decisions that are more repeatable and transparent than other methods that rely on subjective judgement.

## 2.2 Metric requirements – the theory

What form and structure a metric should use is a vexed issue with a range of different approaches having been adopted in different applications. (Maron, Rhodes *et al.* 2013), (Wunder *et al.* 2008) and (Hajkowicz, Collins *et al.* 2009) each review different aspects of metrics employed in conservation tenders. The framework employed by Maron *et al.* (2013) provides an elegant structure for articulating the nature of the benefits that the metric is intended to estimate (Figure 1).

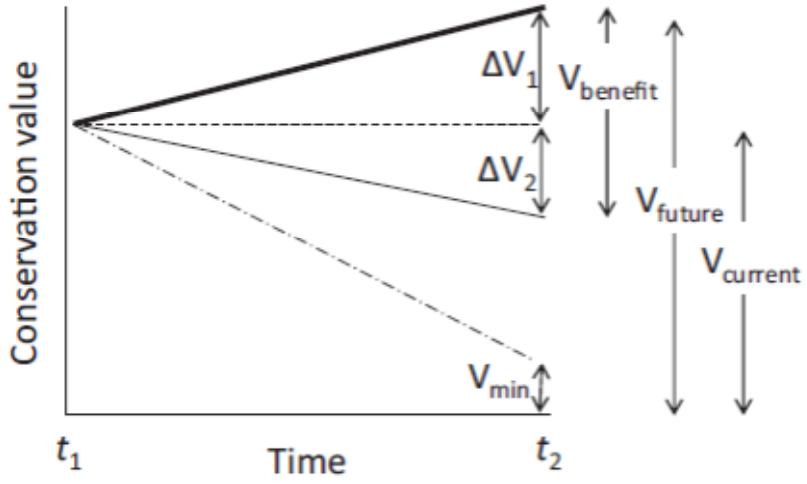


Figure 1: Definitions of change in conservation value (V)

Source: Maron *et al.* (2013), Figure 1. Maron *et al.*(p.2) describe the notation in Figure 1 as follows:

> Change in conservation value (V) at a site between two points in time ($t_1$ and $t_2$) with (bold line) and without (faint line) implementation of a conservation action. The change in the conservation value attributable to the conservation action at time $t_2$ is $V_{benefit}$ (the difference between the two scenarios), but the conservation benefit of the action is often calculated as $V_{future}$ or $V_{current}$. In this example, the conservation action results in both an increase in conservation value (e.g., due to reduced pest density) compared to that at $t_1$ ($\Delta V_1$) and averted los of value ($\Delta V_2$). $V_{min}$ represents a future minimum value that is legally possible within duty of care.

Maron *et al.* (2013) argue that the key issue in metric design for efficient outcomes is additionality or net gain over a realistic baseline assessment of what would occur in the absence of investment. Additionality is often considered a key requirement for economic efficiency because it explicitly requires that the metric measure the new or additional benefits that will result from the change to management. Maron *et al.* (2013) and (Moilanen, Van Teeffelen *et al.* 2009) further urge that consideration to threats to existing condition be taken into account in assessing the future condition baseline from which benefits are assessed. Maron *et al.* (2013) review the metrics of eleven agri-environmental schemes, not all of which were conservation tenders, and find a striking diversity of measures of benefit, none of which were regarded as effective measures of additionality. Similarly Wunder *et al.* (2008) found almost all metrics applied in PES schemes did not attempt to clearly specify additionality.

Despite the argument for a focus on additionality in order to deliver economically efficient investments, Claassen *et al.* (2008), amongst others, note that there may be instances where it is appropriate to reward landholders for past actions for either ethical or behavioural reasons. The ethical justification suggests that these landholders may have foregone income in order to protect the environment and should now be rewarded over their peers who are seeking payment for rehabilitation. The behavioural justification argues that only paying those landholders in a position to deliver improved environmental benefits provides a perverse incentive for them to damage the environment. Another common restriction applies a minimum standard, which becomes the relevant baseline when $V_2$ in Figure 1 falls below that level. The minimum standard generally relates to a duty of care requirement that landholders are expected to deliver in managing their land, and usually relates to property right considerations, sometimes defined as a threshold below which poor land management generates negative externality impacts via soil erosion or other effects.

There is relatively little guidance available in the conservation tender literature as to either the attributes that should be included in any calculation of '*V*' in Figure 1, or the way in which they should be assembled (i.e. the functional structure). The attributes that adequately reflect the investment value are likely to be context and application dependent. While attributes are important, our focus is on the functional structure in which they are combined, and the practical implications of the resulting benefit estimates. In particular we are interested in: 1) comparison of a ratio-scale metric with a non-ratio-scale metric; and 2) employing $V_{future}$ as compared to either $V_{benefit}$ or $V_{benefit}$ given $V_{min}$ for the efficiency of a particular auction mechanism.

## 2.3 Metric functional form – the practice

In order to reflect the current value of an uncertain future benefit the metric should deliver an estimate of the relative expected value from alternative investments. Neither Maron *et al.* (2012) nor others offer much in the way of comment on the functional forms appropriate. Mathematically an expected value would usually require a multiplicative structure employing the probabilities of alternative conservation outcomes scaled to their relative investment values. In practice we observe a wide variety of functional forms employed in conservation tender metrics. Although these metrics often borrow elements from each other they can be distilled into three broad classes:[1]

1. Condition index based metrics with an implied ratio interpretation: metrics employing at their core a scoring system based on vegetation condition scores, commonly with adjustments for landscape context and scarcity. Condition-index based metrics employ an additive form in the condition index component and either an additive (BioMetric -Gibbons, Briggs et al. 2009) or multiplicative (Habitat Hectares - Parkes, Newell et al. 2003) form in

[1] Where outcome oriented metrics are employed they often employ a condition index or MCA oriented approach scaled for the relative value of alternative investments.

translating the condition index to future benefit estimates. The two reflect either substitution or complementarity in the way in which landscape context and other factors are incorporated.

2. Multi-criteria based indices (US CRP - Claassen, Cattaneo et al. 2008; Soil & Water Conservation Society and Environmental Defense Fund 2008), Nature Assist (Hajkowicz *et al.* 2008)). Usually MCA approaches encompass multiple objectives employing at their core a single summation of weighted benefits to yield the benefit estimate and sometimes employing condition indices as subcomponents.

3. Expected value oriented indices (ESP Australia - Whitten *et al.* 2012): metrics which are structured around a set of explicit probabilities describing the likelihood of future events which will impact positively or negatively on the value of an investment.

In the discussion that follows we use the convention of estimating investment values on a per hectare basis, although this need not be the case (see for example Reeson *et al.* 2011 for a discussion of alternative scale orientations of metrics). The scores described above may also be weighted for the degree of future protection of the outcomes of investment (security) and the length of time for which management is contracted in variable length tenders.[2]

Condition indices generally describe the ecological health of site with respect to a reference benchmark. Scores are not usually intended to be relative in the sense that one site could be regarded as half as healthy or only generating half the value to society as another with twice the score, indeed they are often simply spread across the range 0-100. For example, a site in poor health (low condition score) may still be highly valued if, for example, mature trees remain. The challenge where condition scores are employed is in the translation into relative investment values using what is typically a rank-order oriented metric. This is difficult in practice. As Parkes *et al.* (2004) observe:

> Foremost in our minds during the development of the 'habitat hectares approach' was the need to balance the 'science' of a metric for vegetation condition with the need for an approach that was practical and relatively simple.

Parkes *et al.* (2004) also explicitly noted that the economic value of alternative options is outside of the focus of their Habitat Hectares approach, despite the fact that it was subsequently adopted to discriminate between the values of alternative investment options in the BushTender conservation auction.

The translation of relative values into a suitable scoring approach in conservation tenders typically follows either an additive or multiplicative form. In the additive form different elements of ecological condition are weighted according to their contribution to investment values. Measures of new attributes not typically included in condition indices such as landscape context or scarcity, are added to the core condition score. The resultant functional form can be generalised as described in Equation (1).

$$\text{Score} = a \sum w_j B_j \qquad (1)$$

Where: '$w$' are the weights for each of '$j$' benefits '$B$', and '$a$' is area.

The functional form of Equation (1) is identical to the MCA approach described later with the constraint that $B_j$ relate exclusively to biophysical descriptors of the tender site and its landscape context. Depending on which form of $V$ from Figure 1 is intended (usually one of $V_{current}$, $V_{current} + \Delta V_1$ or $V_{benefit}$) $B_j$ will be an estimate of current attribute values, current values adjusted for improvement

[2] These factors are often included directly among the criteria in MCA approaches.

after management change or a differenced estimate of future benefits with and without management or similar.

The typical multiplicative form is shown in Equation (2).

$$\text{Score} = a * V_{current} * V_{change} \quad (2)$$

Where: '$V_{current}$' is the estimate of current conservation value, '$V_{change}$' an estimate of the benefits from changing management, and '$a$' is area.

The functional form of both $V_{current}$ and the benefit estimate is usually as per Equation (1) but with no area weight and with no internal adjustment for management (i.e. $\sum w_j B_j$). The benefit estimate typically is intended to represent either $V_{benefit}$ or $\Delta V_1$ from Figure 1. The main practical difference between the additive and multiplicative approaches is in the assumed benefits from investment. The multiplicative approach estimates service scores as a scalar of the current value of the asset which usually includes a rarity or scarcity weighting. The additive approach often includes a scarcity score as one of the benefits, which is then not differenced in any subsequent step (see for example the detailed description of the BGGW metric in section 3). The assumption in a multiplicative form is that equivalent management impact to a higher initial or current asset value (higher $V_{current}$) is worth more; in the additive form that a high value asset is always worth more. The actual impact of these assumptions depends on the weightings.

The multi-criteria approach (MCA) is intended to 'measure benefit over multiple outcomes' (Hajkowicz *et al.* 2008) although the multiple outcomes could be limited to different aspects of biodiversity as is the case with condition based indices. The approach is mathematically equivalent to that described in Equation 1. However, the MCA approach is an application of cost utility analysis where each of the benefit criteria is assumed to deliver utility to the investor in proportion to their weighting. The compromise programming algorithm used by Hajkowicz *et al.* (2008) for the Queensland NatureAssist Scheme and the MCA basis of the CRP in the US (Soil & Water Conservation Society and Environmental Defense Fund 2008) essentially employs the same logic. Hajkowicz *et al.* (2007) combine their approach with a knapsack optimisation algorithm which is intended to ensure that, subject to the budget limit, the maximum sum of benefits from individual tenders is purchased.

Benefit weights in MCA approaches are intended to reflect the relative preferences of investors for differential investment outcomes. For example NatureAssist weightings were provided by the Queensland Government (Hajkowicz *et al.* 2008). Estimates for the biophysical values within MCA are reliant on either expert judgement or appropriate underpinning models to generate estimates of the benefit in each criterion. For example, a condition index may be used as a proxy measure for one type of $B_j$ in the MCA approach.

Expected value metrics are constructed as an estimate of the expected investment value of a particular investment given the initial starting conservation value, identified threats and damage likelihood to that conservation value, and the management actions applied and their effectiveness in counteracting the identified threats and/or enhancing value independent of threats (Equation 3). This construct facilitates consistent estimation of the relative return on investment from conservation actions offered at $t_2$ in Figure 1.

$$Score = a * f(V_{current}, Th_i * L_i, M_j * E_j) \quad (3)$$

Where: '$V_{current}$' is the estimate of current conservation value , '$L$' are the likelihood of impact for each of '$i$' threats '$Th$', '$E$' are the effectiveness for each of '$j$' management actions available '$M$', and '$a$' is area.

Expected value metrics apply a multiplicative form and could employ either a bounded (typically zero and 100) or unbounded form. When the functional form in Equation (3) is structured to estimate the difference between the score with and without managements offered in a tender the score represents additionality as defined by (Maron, Hobbs *et al.* 2012) inclusive of time and uncertainty as discussed by (Moilanen, Van Teeffelen *et al.* 2009).

Each of the metric forms described above has strengths and weaknesses. Condition index based approaches have the strength of drawing on commonly available established approaches to assessing vegetation condition. However, in neither additive nor multiplicative form are they likely to be ratio-scale comparable. MCA approaches are framed in the well established approach of cost utility analysis but may fail to adequately encompass ecological dynamics in estimating value because of their reliance on estimating final utility values directly. Providing $V_{current}$ is scaled appropriate to the value for differential outcomes an expected value metric is likely to be the most consistent with both economic and ecological theory. However, there is still the difficulty of scaling investment values (faced by all metrics but see Boyd and Banzhaf 2007) and of identifying and estimating the probabilistic impacts of threats and effectiveness of management actions offered. MCA approaches are probably the only viable outcome where multiple ecosystem service values are desired but could be integrated with an expected value approach for each service.

# 3. Does it matter – evidence from a case study

Policy design and implementation always involves some pragmatic tradeoffs. It is clear from Section 2 that most metrics that have been implemented incorporate significant divergences from what would be considered optimal, from an economic efficiency perspective. The question is what are the implications of these tradeoffs for the core objective of the policy? In this section we examine the history and performance of one program, the Australian Government's Environmental Stewardship Program over the period from its inception in 2007 until 2011, encompassing three different metric structures. The ESP provides a unique opportunity to examine the influence of metric design on conservation tender results due to its inclusion of a least one ecological community throughout this period[3] (Box Gum Grassy Woodlands (BGGW) and derived grasslands), and the range of innovations in the metric design.

## 3.1 A brief history of the Environmental Stewardship Program

The Australian Government's Caring for Our Country Environmental Stewardship Program (ESP) has been in operation since 2007. The objective of the project is to 'maintain and improve the quality and extent of targeted high public value assets on private land' (Department of Environment Water Heritage and the Arts and Department of Agriculture Fisheries and Forestry 2008). Environmental Stewardship projects operate through contracts with private land managers from ten to fifteen years. The desired outcomes include:

- Improved habitat across the landscape;
- Increased viability, integrity and buffers to high quality remnants;
- Improvements to the long-term protection of nationally endangered ecological communities;
- Improvements in the condition and function of those ecological communities; and

[3] The ESP was expanded to target four further communities in the 2011 and 2012 rounds but these additional communities are excluded from our analysis.

- Enduring changes to land manager attitudes and behaviours towards environmental protection and sustainable land management practices.

Six funding rounds were implemented as part of the ESP from 2007 to 2011. Three different metrics were applied across the period. The initial metric was derived from the NSW Biometric approach (Gibbons, Briggs *et al.* 2009) and set out in Table 1 (Gibbons and Ryan 2008). The generic Gibbons and Ryan metric functional form for a single site offered in a tender is.[4]

$$CVM = ((R * W_r) + (L * W_l) + ((P_m - P_{um}) | P_{min}) * W_p) * A^x * D * S \quad (4)$$

Where: The parameters are described in Table 1. The subscripts '*m*' and '*um*' describe the managed and unmanaged state of the site respectively and 'min' describes the required duty of care outcome.

Equation (5) is an additive form metric with a condition index at its core. The site score is differenced with a fixed increase in scores following the addition of management, subject to a minimum standard duty of care requirement '$P_{min}$'.

Table 1: Box Gum Grassy Woodlands metric 2008-2009

| | **CVM components** | | | **Weighting** (*W*) | | | | **Scaling** |
|---|---|---|---|---|---|---|---|---|
| Conservation Value Measure or CVM (also known as conservation value score) | Region (*R*) Mean regional rarity of community | | | | 100 | 0.1 | | Partly (99% max score) |
| | Landscape (*L*) | Regional landscape value (10,000 ha) | | 0.25 | 100 | 0.3 | | None |
| | | Landscape value (1000 ha) | | 0.25 | | | | None |
| | | Patch size value (10 ha) | | 0.5 | | | | None |
| | Patch or Site (*P*) (difference between with and without management from current condition) | Component factors | Understorey richness (excluding grasses) | 0.25 | 100 (Calculated as the difference between management and no-management) | 0.6 | CVS out of 100 | Scaled as a proportion of the maximum possible score for each vegetation category |
| | | | Number of mature trees | 0.15 | | | | |
| | | | Length of fallen logs | 0.025 | | | | |
| | | | % native over-storey cover | 0.05 | | | | |
| | | | % native mid-storey cover | 0.025 | | | | |
| | | | % native under-storey cover | 0.1 | | | | |
| | | | % exotic under-storey cover | 0.1 | | | | |
| | | | Over-storey regeneration | 0.1 | | | | |
| | | Interaction terms | Native plant species x Native understorey cover | 0.05 | | | | |
| | | | Mature trees x Regeneration | 0.05 | | | | |
| | | | Mature trees x Logs | 0.05 | | | | |
| | | | Midstorey x Native understorey cover | 0.05 | | | | |
| | Duration (*D*) | | | *D*/15 *100 | | | | None |
| | Security (*S*) | | | 3.7 | | | | None |
| | Area (*A*) | | | 1 | | | | Exponent (*x*) = 0.4 |

Source: compiled from (Gibbons and Ryan 2008).

The Gibbons and Ryan metric differed substantively from the NSW Biometric in the inclusion of a minimum standard duty of care baseline and the application of an exponent to the area term in the equation. The area exponent is an attempt to apply a species area curve relationship in a form which

[4] The score can be aggregated across multiple '*i*' component zones as follows.

$$CVM = \sum_i ((R * W_r) + (L * W_l) + ((P_m - P_{um}) | P_c) * W_p)) * A^x * D * S \quad (5)$$

assumes that the primary basis of the BGGW program is to conserve the greatest number of species (consistent with Stephens *et al.* 2002).

A number of aspects of the BGGW Project were reviewed during 2009, including the BGGW CVM (or metric) (Whitten *et al.* 2009). Following the review two significant changes were made to the metric, along with a range of minor changes (the latter being largely for the purposes of consistency and ease of implementation). The major changes were: (1) the removal of the area exponent which had the effect of a strong preference for small sites over large and inadequate emphasis on the goal of increasing the extent of the listed community;[5] and (2) a halving of the security weighting to reflect the economic cost savings that could be attributed to a permanently protected site when compared to a site that required a repeat contract. The revised box gum grassy woodlands metric remains an additive form, condition index based metric but reflects a significant difference in preferences for larger sites and permanent protection.

The Whitten *et al.* (2009) review also identified that, in common with the majority of approaches, risk and uncertainty were not explicitly addressed in the Gibbons and Ryan approach, and that the approach was not fully compatible with the ecological state and transition approach being adopted by the ESP (Australian Government Department of the Environment and Heritage 2006). As a result a revised metric was constructed for the 2011 and 2012 rounds of the ESP, known as the multiple ecological community (MEC) metric (which extended the metric to other ecological communities) (Whitten, Zammit *et al.* 2012).[6]

$$\textit{Conservation Value Score} = site_{15} * a * D * S \qquad (6)$$

And $site_{15} = f(site_I, P_{loss}, P_{gain})$

Where: '$site_{15}$' is expected final score per hectare for a site in 15 years time; '$site_I$' initial site score per hectare assessed as a state and transition model state; '$P_{loss/gain}$' are probabilities of degradation/improvement to a one step lower/higher ecological state and transition model state, '$a$' is area, $D$ and $S$ are duration of contract and security.

Equation (6) is an expected value index with probabilities of loss and gain calculated from the presence or absence of threats and management actions and their likely impact on the ecological state of the site (a discretised indicator of condition, and the basis of state and transition models of ecological change). The '$site_I$' values are scaled to the relative investment values for different ecological condition states. As such the structural form provides a metric that is consistent with the ecological dynamics of Box Gum Grassy Woodland Communities (represented in the state and transition model), and consistent with the economic prioritisation needs of a ratio-scale comparable metric.

## 3.2 Methodology

Two different metric functional forms, with a further major variation within the additive form, have been applied to the ESP program over the past five years as the Australian Government has sought to improve investment decisions. From a policy perspective the key question is the extent to which

---

[5] The species area curve was used in the Gibbons and Ryan metric to score an individual site. However, the original intent of inclusion of species area curves in metrics is to represent the diminishing increase in species observed when adding unconnected sites into a conservation set. In the context of a single previously connected ecological community the use at site scale is incompatible with ecological theory.

[6] The MEC metric score can be aggregated across multiple '$i$' component zones as follows.

$$\textit{Conservation Value Score} = [\sum_{i=1-n} (site_{i,15} * a_{sitei})] * D * S \qquad (7)$$

Where the additional subscript $i$ denotes the relevant PMU ($i$ = 1, … , $n$).

these different ways of measuring the relative investment value's impacted on investment decisions. In order to assess the impact we compare different metric forms and different investment objectives across bid data from successive rounds of the ESP. We are interested in the impacts of two different questions. First, what is the impact of different benefit estimates within a single metric functional form on the likely set of investments? Second, what is the impact of a different metric functional form on the likely set of investments?

In order to test these two questions on real bid data we are able to compare forms as set out in Table 2. Tests of different benefit estimates are represented by rows in Table 2. For example, we compare the set of bids that are likely to be accepted for three different benefit estimates for the MEC metric data set (2011 bid data only). We can similarly compare modifications of the Gibbons and Ryan 2008-2009 metric and the revised 2010 metric, and finally a modified MEC metric across the entire data set. Tests of different metric functional form are able to be performed where ticks appear vertically in columns with the exception of the MEC data set in isolation. Thus we can compare the 2008-09 and 2010 metric versions, as well as a modified MEC data set across all years.

Table 2: Metric benefit forms compared

| **Metric compared** | **Alternative benefit estimates (from Figure 1)** | | |
|---|---|---|---|
| | ***Total future benefits ($V_{future}$)*** | ***Additionality benefit ($V_{benefit}$)*** | ***Additionality over duty of care*** |
| MEC data set (2011) | ✓ | ✓ | ✓ |
| BGGW metric version 1 and 2 (2008-10) | ✓ | ✓ | ✓ |
| Modified MEC metric across a combined data set (2008-2011) | ✓ | ✓ | |

There are several methods by which we can compare outcomes. First, we can compare the degree to which the set of recommended bids differs across both benefit estimates and functional forms. Second, we can examine the discriminatory power of the metric, always keeping in mind that difference benefit estimates reflect differing assumptions about the value derived from investment. Finally, we can examine whether there are significant divergences in the effectiveness of metrics in purchasing desired outcomes. We note that the analysis is in a relatively early and qualitative stage and we have not as yet been able to identify appropriate statistical measures of difference (suggestions welcome).

# 4. Results

At this point we report results for two types of comparison from Table 2 that have been undertaken to date. For the MEC metric (2011 data) we are able to provide a set of comparisons across different benefit estimates ($V_{future}$, $V_{benefit}$ and $V_{future}$ less $V_{min}$). Second, we provide some comparative assessment of the original and modified BGGW metrics (both formulated as $V_{future}$ less $V_{min}$). Work is progressing on difference benefit estimates for the two forms of BGGW metric (Table 1) and as well as on a modified version of the MEC metric that can be implemented across a combined data set from which we can directly compare the expected value and condition index forms.

## 4.1 Comparison of different benefit estimates

First we compare the degree to which the three benefit estimates differ in the set of bids recommended for contracts for the 2011 MEC round of the Environmental Stewardship Program. Figure 2 shows the degree of alignment across the three benefit indices as a potential budget is progressively exhausted. Four different relationships can be seen across the potential range of budgets. At very low budget levels there is complete agreement in the set of bids recommended for

contracts. As budgets increase significant disagreement emerges across a budget of around $7m to $25m, with the set of bids recommended by all measures falling well below 50% and around 20% of bids are only recommended by one measure. Between budgets of $25m and $50m there is substantive agreement in the set of bids (around 85 to 90% are recommended by all three) before very large budgets make metrics largely redundant (>$50m). In part the third and fourth phases are driven by the diminishing number of bids available to choose between as the number of bids chosen increases. An alternate way of considering the differences in benefit estimates is shown in Table 3 for three hypothetical budgets. The typical ESP contract offer rate within the data is between budgets of $42m and $48m by which time there is quite high agreement between benefit measures, but there still remains a total variation in the number of contracts offered of seven, and all three benefit estimates are in agreement for less than 80% of bid value.

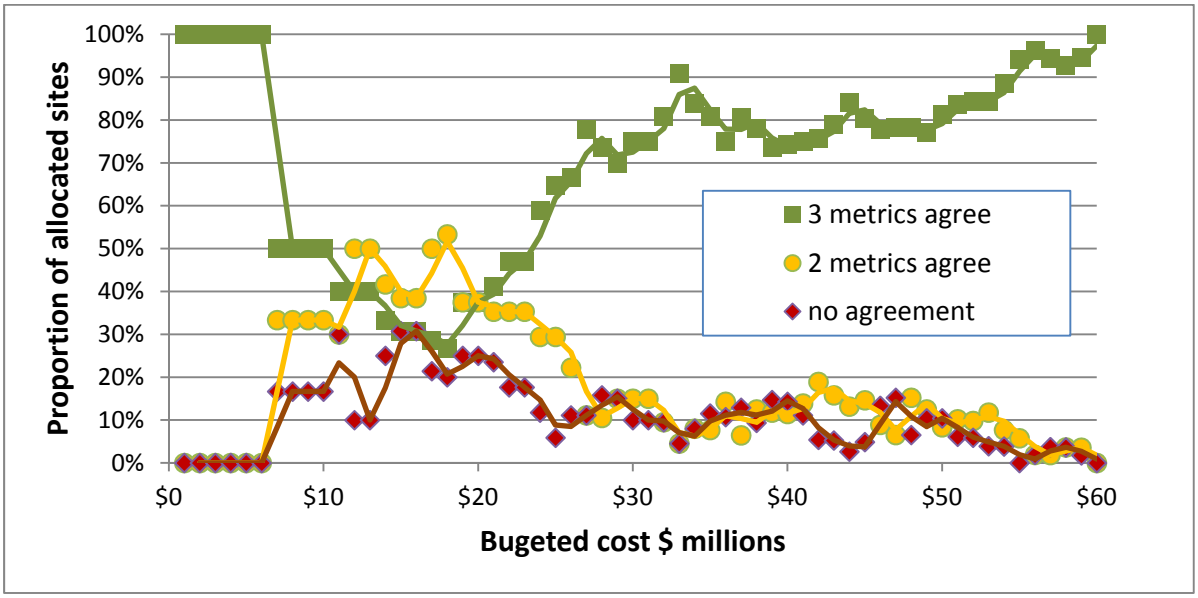


Figure 2: Degree to which three benefit estimates differ in MEC metric

Table 3: Summary of MEC metric comparisons for different budgets

| **Comparison measure** | **Benefit estimate** | **Alternative budgets** | | |
|---|---|---|---|---|
| | | ***$30m*** | ***$42m*** | ***$48m*** |
| **Percentage of value allocated** | Total benefits metric | 77% | 92% | 95% |
| | Additionality metric | 77% | 92% | 96% |
| | Additionality over duty of care | 78% | 92% | 95% |
| **Proportion of projects allocated** | Total benefits metric | 30% | 61% | 79% |
| | Additionality metric | 30% | 61% | 77% |
| | Additionality over duty of care | 34% | 57% | 68% |
| **Number of projects allocated** | Total benefits metric | 17 | 34 | 44 |
| | Additionality metric | 17 | 34 | 43 |
| | Additionality over duty of care | 19 | 32 | 38 |
| **Proportional of allocation agreement of bid value** | 3 benefit estimate agree | 75% | 76% | 78% |
| | 2 benefit estimates agree | 15% | 19% | 15% |
| | No benefit estimates agree | 10% | 5% | 7% |

Two alternate ways of viewing the discriminating power of different benefit estimates are shown in Figure 3 and Figure 4. In Figure 3 the distribution of allocation scores as dollars per normalized score is shown. Scores are normalized because the different benefit estimates results in different score ranges. As would be expected the spread of scores for a total benefit ($V_{current}$) metric is higher than either benefit variant. This can be interpreted as the total benefit metric having greater

discriminatory power between bids (assuming that the benefit is appropriately defined). An alternative efficiency measure, the degree to which value is added per budget allocated is shown in Figure 4. Though there is relatively little difference between the benefit estimates, their order does swap and additionality subject to duty of care appears marginally the most efficient in the investment range which the ESP program typically operates.

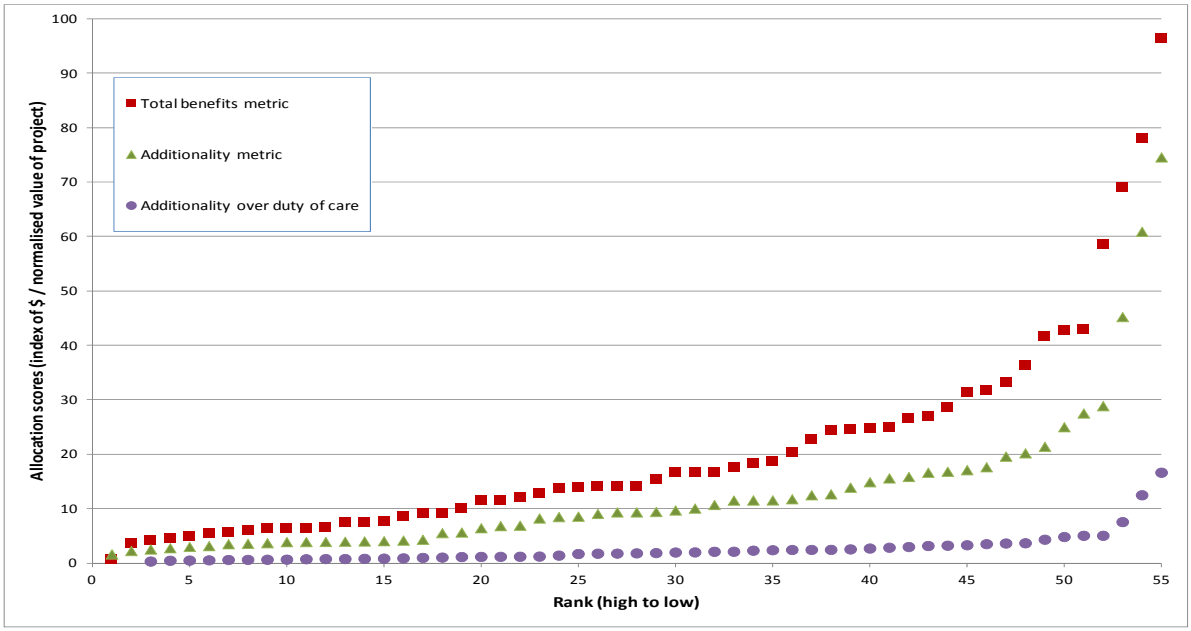


Figure 3: Distribution of allocation scores for MEC metric variants

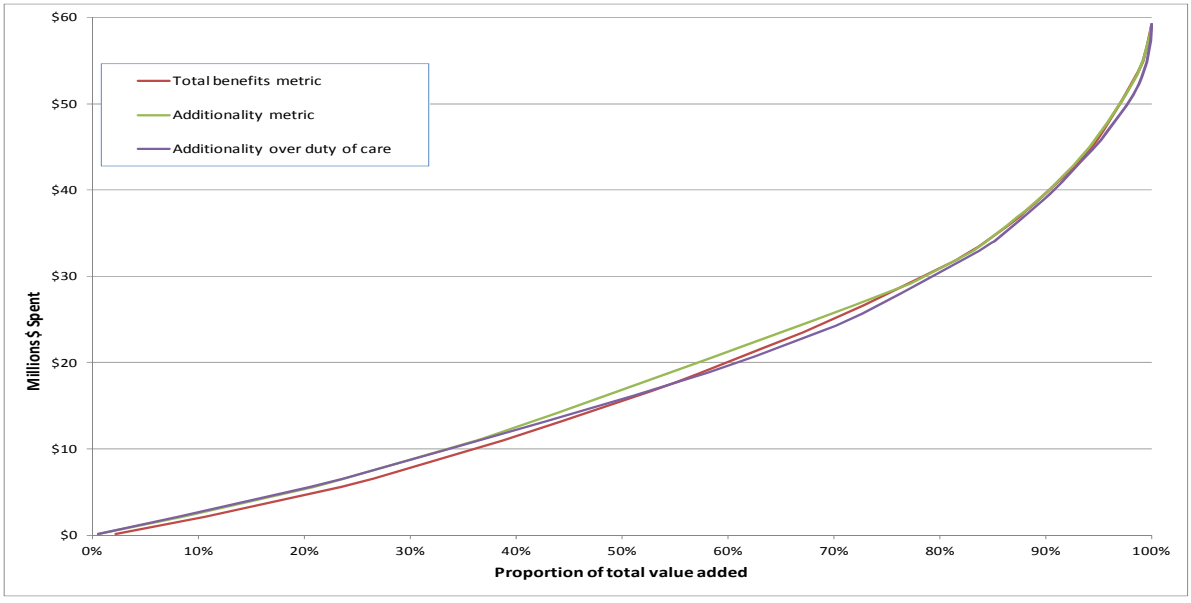


Figure 4: Relative efficiency in normalised purchasing value for MEC metric variants

## 4.2 Comparison of different metric functional forms

A similar analysis of two metric functional forms, Gibbons and Ryan and the revised Box Gum Grassy Woodlands (subsequently referred to as the revised BGGW), across the 2007-2009 rounds of the Environmental Stewardship Program has been undertaken. Both metrics can be described as $V_{future}$ less $V_{min}$. The degree of agreement across the two metrics is shown in Figure 5. Despite the similarity in functional form there is considerable disagreement between the two metrics at very low budgets

(less than $5m). Between an investment cost of $5m and around $17m the level of agreement stabilizes in a band between 70% and 85% before the diminishing number of bids available to choose between as the number of bids chosen reduces budget disagreement beyond $17m. The consequences of the divergence between metrics for a suite of budget scenarios are shown in Table 4 (the typical ESP offer rate is between the $5.5m and $12m bands). The divergence between metrics remains as high as 20% of the budget even when approximately 79% of bids are offered contracts.

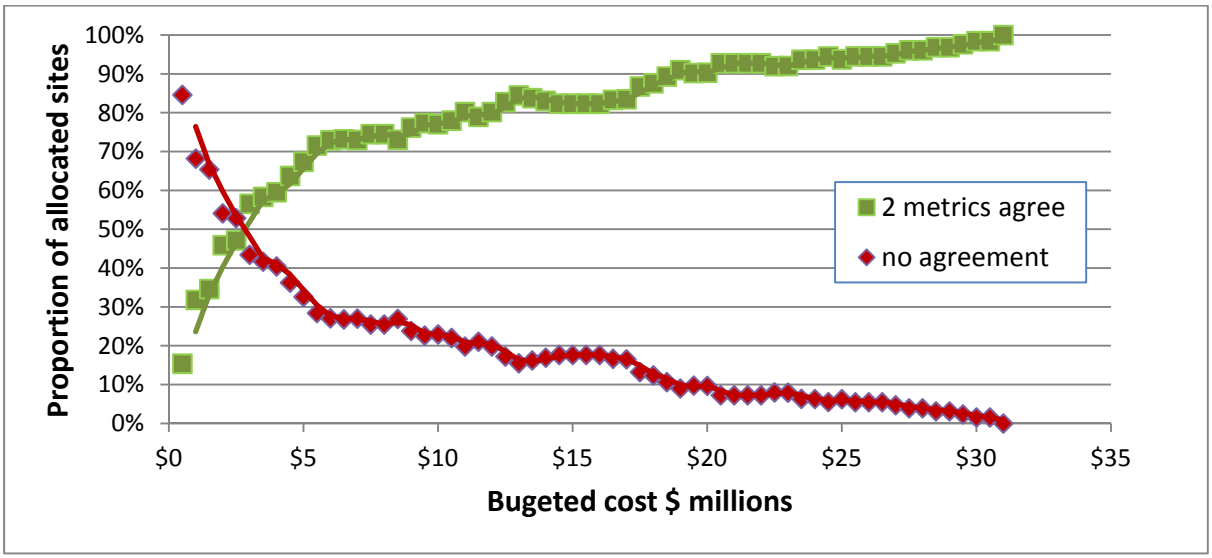


Figure 5: Degree to which Gibbons and Ryan and revised BGGW agree

Table 4: Summary of Gibbons and Ryan and revised BGGW comparisons

| **Comparison measure** | **Benefit estimate** | **Alternative budgets** | | |
|---|---|---|---|---|
| | | ***$3m*** | ***$5.5m*** | ***$12m*** |
| **Percentage of value allocated** | Gibbons and Ryan | 83% | 89% | 95% |
| | Revised BGGW | 65% | 77% | 92% |
| **Proportion of projects allocated** | Gibbons and Ryan | 45% | 63% | 79% |
| | Revised BGGW | 48% | 64% | 84% |
| **Number of projects allocated** | Gibbons and Ryan | 57 | 81 | 101 |
| | Revised BGGW | 62 | 82 | 108 |
| **Proportional of allocation agreement of bid value** | 2 benefit estimate agree | 57% | 72% | 80% |
| | No benefit estimates agree | 43% | 28% | 20% |

Figure 6 and Figure 7 show two alternate ways of viewing the discriminating power for the Gibbons and Ryan and revised BGGW. The distribution of scores as dollars per normalized score is very similar for the two metric forms indicating that the discriminatory power of the two metrics across bids is similar with some relatively minor divergence in the tail of low ranked bids (beyond 110). The alternative efficiency measure, the degree to which value is added per budget allocated is shown in Figure 7. Despite the similarity in discriminatory power, the revised BGGW metric appears to be much more efficient within and beyond the investment range which the ESP program typically operates. That is, a much higher proportion of project value is added at a lower budget than through the Gibbons and Ryan metric. This conclusion emphasizes the potential that seemingly small changes to metrics can have large impacts on the ability to discriminate between the relative investment value of alternative bids.

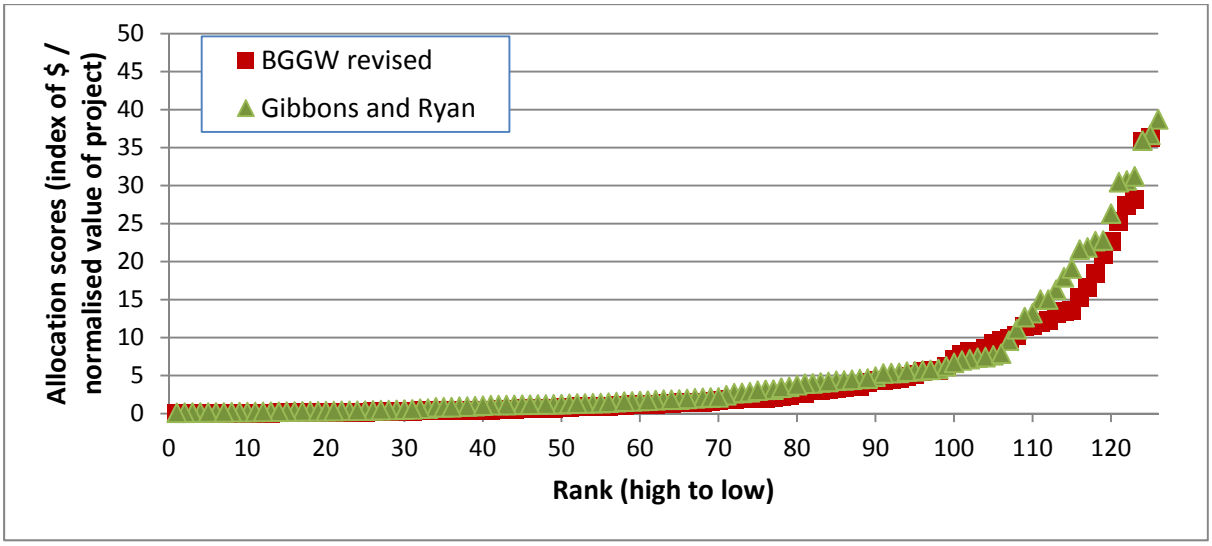


Figure 6: Distribution of allocation scores for Gibbons and Ryan and revised BGGW

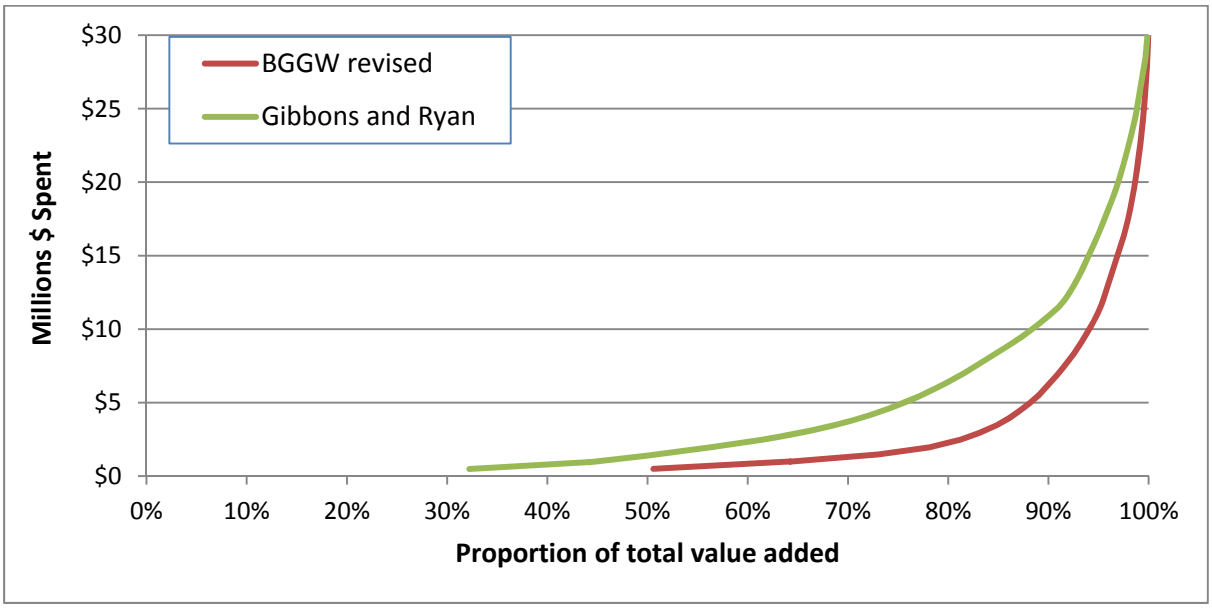


Figure 7: Relative efficiency in normalised purchasing value for MEC metric variants

## 5. Discussion and conclusions

A wide variety of both metric functional forms and benefit estimates have been applied in conservation tenders to date with relatively little critical discussion and analysis of the consequences for the bundles of conservation benefits selected for investment. In this paper we have explored a number of elements of the design and performance of metrics in conservation tenders using theory and practice. We conclude that ideally a metric should meet a number of practical and theoretical performance criteria for use in a conservation auction or similar approach involving payment for relative environmental services. These performance criteria are: a scientifically robust measure of the biophysical interrelationships response to different investments; consistency in repeat applications; transferability and replicability across the desired application domain; economically and socially defensible for the investment values generated; and ratio-scale comparability to support decisions about relative funding levels. Theory further suggests that the estimate should represent the net additional benefit from investment over the option of not investing (additionality) taking into

account the threats and consequences of not acting and the likelihood of investment being successful in delivering the desired environmental outcome.

An important question for policy makers is the extent of impact of the functional form of the benefit metric and the nature of the benefit estimated on which tenders are selected for investment. That is, does the difference in functional structures or the difference benefit estimates have a strong or weak impact on investment decisions by governments? If weak then there is much greater flexibility available in metric structure so long as it consistently discriminates across the tenders submitted. Furthermore, a relatively weak relationship indicates greater freedom to include other elements besides pure economic efficiency into metrics with relatively low risk of distorting decision preferences. For example, if there is little evidence of a significant difference in the recommended investment bundle across $V_{future}$, $V_{benefit}$ and $V_{future}$ less $V_{min}$ estimates of benefit there is greater latitude to actively support those who have historically managed environmental assets well. If strong, then policy makers will need to invest greater effort to ensure that a metric under consideration meets performance requirements.

In this paper we report preliminary tests across the three metrics applied in the Environmental Stewardship Program in Australia from 2007-2012. We compare three benefit estimates, $V_{future}$, $V_{benefit}$ and $V_{future}$ less $V_{min}$, for the MEC metric bid data from 2011. We also compare the Gibbons and Ryan and revised BGGW, two different functional forms of metric although both are condition index based additive metrics, across the 2007-2009 data. Research is proceeding to undertake a similar and more comprehensive analysis across bid data for the full period applying different benefit estimates and the three forms of metric. Note that not all of the desired performance criteria are directly testable using bid data. Our results are therefore directed towards the influence of benefit estimates, particularly divergence from additionality, and the functional structure of metrics (which relates to scientific robustness, ratio-scale comparability and impact of uncertainty).

Our preliminary results suggest that the difference in benefit estimates will only have a large impact on the set of bids recommended at smaller budgets relative to the total of those offered. For the ESP program, where acceptance rates have typically fallen between 60% and 80% of bids submitted, the difference in form of benefit estimates for the MEC metric version suggests little difference in the allocative efficiency or total benefits delivered. There is some divergence in the set of bids recommended in this range but the consequences for the estimated return from investment are relatively small. We suggest that the expected value MEC metric is relatively robust under different benefit formulations. A similar analysis comparing two different functional forms of metric, although each is a condition index, additive based metric, suggests greater divergence in performance, and in particular a significantly larger impact on the conservation value delivered. Our conclusion in this instance is that it is important to align the functional structure to the desired investment outcomes. Indeed a key reason why innovation in metric design was pursued in the Environmental Stewardship Program was that the metric structure only partially reflected the investment priorities of government.

We note that there are a number of other elements of economic theory that have not been addressed in the research in this paper. In particular we have not as yet addressed the extent to which landholders are able to exert information rents about the relative conservation values offered by their bids in conservation auctions (Connor, Ward *et al.* 2008; Juutinen *et al.* 2013). This literature is important to establishing the fundamental underpinnings of economic efficiency in conservation tenders and the extent to which landholder pricing decisions reflect their costs. If information rents can be exercised differentially in different benefit forms or through different functional forms there is the potential for the preliminary results set out in this paper to be confounded. However, given that the metric functional form, details around the form of benefit estimate, and comparison of environmental scores was not available in the Environmental Stewardship Program it seems unlikely

that landholders would be able to exercise information rents differentially across the different auction settings explored in this paper.

## References


Australian Government Department of the Environment and Heritage (2006). Environmental Protection and Biodiversity Conservation Policy Statement: White Box - Yellow Box - Blakelys Red Gum Grassy Woodlands and derived native grasslands. Australian Government Department of the Environment and Heritage.

Babcock, B.A., Lakshminarayan, P.G., Wu, J. and Zilberman, D. (1997). Targeting tools for the purchase of environmental amenities, *Land Economics* 73, 325-339.

Boyd, J. and Banzhaf, S. (2007). What are ecosystem services? The need for standardized environmental accounting units, *Ecological Economics* 63, 616-626.

Cattaneo, A., Hellerstein, D., Nickerson, C. and Myers, C. (2006). Balancing the Multiple Objectives of Conservation Programs. Economic Research Service, USDA, Washington DC.

Claassen, R., Cattaneo, A. and Johansson, R. (2008). Cost-effective design of agri-environmental payment programs: U.S. experience in theory and practice, *Ecological Economics* 65, 737-752.

Connor, J.D., Ward, J.R. and Bryan, B. (2008). Exploring the cost effectiveness of land conservation auctions and payment policies*, *Australian Journal of Agricultural and Resource Economics* 52, 303-319.

Cook, C.N., Wardell-Johnson, G., Keatley, M., Gowans, S.A., Gibson, M.S., Westbrooke, M.E. and Marshall, D.J. (2010). Is what you see what you get? Visual vs. measured assessments of vegetation condition, *Journal of Applied Ecology* 47, 650-661.

Department of Environment Water Heritage and the Arts and Department of Agriculture Fisheries and Forestry (2008). Environmental Stewardship: Strategic Framework. Caring for Our Country.

Ebert, U. and Welsch, H. (2004). Meaningful environmental indices: a social choice approach, *Journal of Environmental Economics and Management* 47, 270-283.

Feather, P. and Hellerstein, D. (1997). Calibrating benefit function transfer to assess the conservation reserve program., *American Journal of Agricultural Economics* 79, 151-162.

Gibbons, P., Briggs, S.V., Ayers, D., Seddon, J., Doyle, S., Cosier, P., McElhinny, C., Pelly, V. and Roberts, K. (2009). An operational method to assess impacts of land clearing on terrestrial biodiversity, *Ecological Indicators* 9, 26-40.

Gibbons, P. and Ryan, P. (2008). A conservation value index for Box Gum Grassy Woodland. Final Report to Accompany Version 1.6. Fenner School of Environment and Society. The Australian National University, Canberra.

Gorrod, E.J. and Keith, D.A. (2009). Observer variation in field assessments of vegetation condition: Implications for biodiversity conservation, *Ecological Management & Restoration* 10, 31-40.

Hajkowicz, S., Collins, K. and Cattaneo, A. (2009). Review of Agri-Environment Indexes and Stewardship Payments, *Environmental Management* 43, 221-236.

Hajkowicz, S., Higgins, A., Miller, C. and Marinoni, O. (2008). Targeting conservation payments to achieve multiple outcomes, *Biological Conservation* 141, 2368-2375.

Hajkowicz, S., Higgins, A., Williams, K., Faith, D.P. and Burton, M. (2007). Optimisation and the selection of conservation contracts*, *Australian Journal of Agricultural and Resource Economics* 51, 39-56.

Juutinen, A., Mäntymaa, E. and Ollikainen, M. (2013). Landowners' conservation motives and the size of information rents in environmental bidding systems, *Journal of Forest Economics* 19, 128-148.

Maron, M., Hobbs, R.J., Moilanen, A., Matthews, J.W., Christie, K., Gardner, T.A., Keith, D.A., Lindenmayer, D.B. and McAlpine, C.A. (2012). Faustian bargains? Restoration realities in the context of biodiversity offset policies, *Biological Conservation* 155, 141-148.

Maron, M., Rhodes, J.R. and Gibbons, P. (2013). Calculating the benefit of conservation actions, *Conservation Letters*n/a-n/a.

McCarthy, M.A., Parris, K.M., Van Der Ree, R., McDonnell, M.J., Burgman, M.A., Williams, N.S.G., McLean, N., Harper, M.J., Meyer, R., Hahs, A. and Coates, T. (2004). The habitat hectares approach to vegetation assessment: An evaluation and suggestions for improvement, *Ecological Management & Restoration* 5, 24-27.

Moilanen, A., Van Teeffelen, A.J.A., Ben-Haim, Y. and Ferrier, S. (2009). How Much Compensation is Enough? A Framework for Incorporating Uncertainty and Time Discounting When Calculating Offset Ratios for Impacted Habitat, *Restoration Ecology* 17, 470-478.

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

Reeson, A.F., Rodriguez, L.C., Whitten, S.M., Williams, K., Nolles, K., Windle, J. and Rolfe, J. (2011). Adapting auctions for the provision of ecosystem services at the landscape scale, *Ecological Economics* 70, 1621-1627.

Ribaudo, M.O., Hoag, D.L., Smith, M.E. and Heimlich, R. (2001). Environmental indices and the politics of the Conservation Reserve Program, *Ecological Indicators* 1, 11-20.

Rolfe, J. and Windle, J. (2011). Comparing a best management practice scorecard with an auction metric to select proposals in a water quality tender, *Land Use Policy* 28, 175-184.

Soil & Water Conservation Society and Environmental Defense Fund (2008). Conservation Reserve Program (CRP) Program Assessment. Soil & Water Conservation Society and Environmental Defense Fund.

Stephens, T., Brown, D. and Thornley, N. (2002). Measuring Conservation Achievement: concepts and their application over the Twizel Area. Science for Conservation, New Zealand Department of Conservation, Wellington.

Stoneham, G., Chaudri, V., Ha, A. and Strappazon, L. (2003). Auctions for conservation contracts: an empirical examination of Victoria's Bush Tender Trial, *The Australian Journal of Agricultural and Resource Economics* 47, 477-500.

Whitten, S.M., Doerr, E., Doerr, V., Langston, A. and Wood, A. (2010). Multiple Ecological Communities Conservation Value Metric. A report for the Australian Government Department of Environment, Water, Heritage and the Arts, Canberra.

Whitten, S.M., Gorddard, R., Langston, A. and Reeson, A. (2009). A Review of the Box Gum Grassy Woodlands Environmental Stewardship Project Metric. A report for The Australian Government Department of the Environment, Water, Heritage and the Arts, Canberra.

Whitten, S.M., Zammit, Z., Langston, A., Doerr, V., Burns, E., Doerr, E. and Attwood, S. (2012). Money for nothing or payments for biodiversity? Design and performance of the Australian Government's Environmental Stewardship Program conservation tender metric, *BioEcon*.

Wu, J., Zilberman, D. and Babcock, B.A. (2001). Environmental and Distributional Impacts of Conservation Targeting Strategies, *Journal of Environmental Economics and Management* 41, 333-350.

Wunder, S., Engel, S. and Pagiola, S. (2008). Taking stock: a comparative analysis of payments for environmental services programs in developed and developing countries, *Ecological Economics* 65, 834-852.

Zabel, A. and Roe, B. (2009). Optimal design of pro-conservation incentives, *Ecological Economics* 69, 126-134.