Open access • Journal Article • DOI:10.1111/J.1472-4642.2012.00947.X

To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale — Source link

Ayesha I. T. Tulloch, Karen Mustin, Hugh P. Possingham, Judit K. Szabo ...+1 more authors

Institutions: Australian Research Council, Charles Darwin University

Published on: 01 Apr 2013 - Diversity and Distributions (Wiley-Blackwell)

Topics: Protected area

To boldly go where no volunteer has gone before: predicting volunteer activity to prioritise surveys at the landscape scale

Ayesha I.T. Tulloch[1], Karen Mustin[1], Hugh P. Possingham[1], Judit K. Szabo[2] and Kerrie A. Wilson[1]

1. Australian Research Council Centre of Excellence for Environmental Decisions, School of Biological Sciences, University of Queensland, St Lucia QLD 4072, Australia

2. Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT 0909, Australia

Running title: Predicting volunteer activity

ABSTRACT

Aim

To identify relationships between volunteer bird survey effort and motivations in order to prioritise investment in future surveying activities.

Location

South west Western Australia, a global biodiversity hotspot.

Methods

We developed nine hypotheses for volunteer motivations to predict the probability of a bird survey being undertaken anywhere in the landscape using data from the New Atlas of Australian Birds. We then established three goals for surveying in the study region: 1) equal representation of surveys across the landscape, 2) surveys stratified by habitat type, and 3) representation of surveys in protected areas. We developed a function to estimate the benefit of investing in professional surveys, given the probability of a volunteer survey taking place and the survey goal, and calculated the cost of meeting a surveying goal with and without accounting for the probability of cells not being surveyed by volunteers.

Results

A model combining the location of protected areas, location of previous records of threatened species, and habitat diversity, was the strongest predictor of the probability of a volunteer bird survey being conducted. Each surveying goal resulted in different areas being prioritised for future surveying, indicating the importance of setting clear objectives before undertaking broad-scale monitoring or surveying activities. If our primary goal is stratified protected area representation in surveys, there are huge cost savings if only protected areas with a 70% predicted probability of not being surveyed by volunteers were selected for professional surveys.

Main conclusions

Professional sampling in survey gaps is required to reduce bias in volunteer-collected datasets. Using models of volunteer behaviour we can identify areas unlikely to be surveyed. If these areas are important for the project objective, then we can either provide incentives for volunteers or carry out professional surveying. These analyses are best done before data collection commences.

Keywords: biological atlas, citizen science, conservation planning, resource allocation, species distribution modelling, volunteer monitoring

(A)INTRODUCTION

Biodiversity is declining worldwide, and in order to prioritise actions to mitigate threats we need to evaluate the current status and distributions of the species we are trying to recover and protect (Walsh *et al.* 2012). Surveys across large scales (using a repeatable sampling method to estimate the number of individuals or the diversity of species) allow us to make such prioritisations. With on-going monitoring (the systematic acquisition of data over time) we can keep track of how well our recovery efforts are going. Atlas projects are broadly defined as landscape-level collections of spatially explicit data on species occurrences (Osborne & Tigar 1992), most often contributed by volunteers (Dunn & Weston 2008; Robertson *et al.* 2010). Volunteer-collected or 'citizen science' monitoring databases have now been established in many parts of the world representing both marine and terrestrial taxa (e.g. Battersby & Greenwood 2004; Schmeller *et al.* 2009; Silvertown 2009). The cost-savings from enlisting the support of the volunteering public in such projects are substantial. For example, in the United Kingdom it has been estimated that volunteers contribute 1.6 million hours annually to bird surveys, work that would cost over US$30 million per annum if undertaken by professionals (Danielsen *et al.* 2009).

Globally, atlases are crucial for engaging people in monitoring and conservation as well as answering questions related to conservation, management and theoretical ecology (Underhill *et al.* 1991; Donald & Fuller 1998; Dunn & Weston 2008; Pomeroy *et al.* 2008; Robertson *et al.* 2010), but issues remain with data gaps and biases (Romo *et al.* 2006; Boakes *et al.* 2010; Botts *et al.* 2010), and with maintaining volunteer interest and objectivity (Booth *et al.* 2011). Species occurrence data, such as those compiled during volunteer atlases often exhibit strong spatial and temporal biases in survey effort (Osborne & Tigar 1992; Romo *et al.* 2006; Boakes *et al.* 2010), meaning that some places are more likely to be surveyed than others, and surveys will not be evenly distributed in time (Bas *et al.* 2008; Sparks *et al.* 2008; Phillips *et al.* 2009). Other problems to be dealt with in volunteer-collected datasets include observer error and heterogeneity in the ability of observers to detect species (Kery *et al.* 2006; Etterson *et al.* 2009).Without robust and unbiased monitoring systems (Yoccoz *et al.* 2001), inferences on species' habitat preferences and distribution have more omission and commission errors (Barry & Elith 2006; Rondinini *et al.* 2006; Boakes *et al.* 2010), which can affect the reliability of models or conservation prioritisation analyses using that data (Rondinini *et al.* 2006; Moilanen *et al.* 2009).

In order to improve the quality of atlas data, under-represented regions can be targeted by professional surveys or by encouraging and directing volunteers, or the sampling design adjusted to sample the environmental variation across the landscape (Table 1). Continued data collection incurs a cost in both resources and time (Hauser *et al.* 2006; Grantham *et al.* 2009), and due to limited budgets, planners of monitoring programs or users of the data have to prioritise future efforts. There is now a large body of literature that explores prioritising or optimising monitoring, including the selection of suitable indicator species (Fleishman & Murphy 2009; Tulloch *et al.* 2011), threatened taxa (Regan *et al.* 2008), comparing monitoring and survey protocols (e.g. Munson *et al.* 2010), improving survey methodology (Joseph *et al.* 2006; Rhodes *et al.* 2006a), and integrating cost in survey design (e.g. Carlson & Schmiegelow 2002;

Gerber *et al.* 2005; Loyola *et al.* 2009; Tulloch *et al.* 2011). Understanding the factors that influence the spatial distribution of volunteer survey effort, such as the motivation for volunteers to survey particular areas, is the key to predicting where gaps are likely to occur in future surveys, and thereby prioritising future surveying efforts. Unfortunately, information on volunteer motivations is not routinely collected in citizen science datasets, and owing to the complex nature of human behaviour these types of data are difficult to infer. One way to make use of limited data is to use species distribution models, which can be parameterised with few variables. While priority areas for future surveys have been derived through species distribution models of target species (Rachlow & Svancara 2006; Rhodes *et al.* 2006b; Jones 2011), the likely sampling patterns of the volunteers who undertake surveys have not been explored.

Substantial worldwide volunteer effort has been directed at bird atlases (over 300 atlases engaging over 100,000 participants; Dunn & Weston 2008) due to the popularity of bird watching and the ease of conducting bird surveys relative to other species (e.g. bats, marine invertebrates) that often require special monitoring equipment or expertise. The New Atlas of Australian Birds (Barrett *et al.* 2003) is a database that relies almost entirely on volunteers. It forms the basis for national bird population estimates and provides data to inform conservation assessments by environmental consultants and policy and planning by local and state governments (Garnett *et al.* 2011; Szabo *et al.* 2011). We chose this database as one of its key aims is to involve the community in the conservation and monitoring of birds (Barrett *et al.* 2003). From 1997 to 2010 more than 7000 surveyors have contributed over 500,000 surveys, resulting in more than 8.5 million records. However, recent analyses of bird atlas data in Australia and elsewhere have found biases in the species surveyed by volunteers (Booth *et al.* 2011; Tulloch & Szabo accepted) and high spatio-temporal variability in sampling effort (Szabo *et al.* 2007), which has implications for applying the data to conservation or management objectives especially in under-sampled or remote areas (Yoccoz *et al.* 2001). To fill knowledge gaps, we need to better understand the current and potential distribution of on-going survey effort, to address key data deficiencies in order to increase the usefulness of atlas data for monitoring and scientific research.

Limited funding for conservation worldwide means that decisions need to be made about where and how to fill gaps, and who will be collecting survey data. We present an approach that enables atlas coordinators or users to prioritise allocation of funding to address data deficiencies, by using knowledge of the motivations and biases of the data collectors to predict their future actions. Establishing explicit protocols to take into account the motivations of volunteers to conduct surveys in particular areas will allow more efficient investment in professional surveys or volunteer incentives to achieve the desired targets. The specific objectives of our study are to:

1. use the geographic location of past surveys to predict the future distribution of volunteer survey effort,

2. assuming that volunteer effort is spatially predictable, extract and apply models of motivational factors, to predict the probability of areas being surveyed in the future, and

3. account for the probability of an area being surveyed by volunteers to help prioritise future investment in professional surveys.

(A)METHODS

(B)Study area

Our study focused on an extensive area (356,717 km^2) of south-west Western Australia (WA), encompassing a biodiversity hotspot as defined by Conservation International (Mittermeier *et al.* 1998; Myers *et al.* 2000). The region is of conservation significance due to high plant endemism (53% of 5571 species) as well as the removal of approximately 70% of native vegetation for cropping and grazing in the last 100 years (Saunders *et al.* 1991; Saunders *et al.* 1993; Hopper & Gioia 2004), which has led to declines and local extinctions of flora and fauna. For instance, 57 of the 280 native bird taxa in the region are now of conservation concern (Mittermeier *et al.* 1998; Garnett *et al.* 2011).

The extent of the study area was converted to a geographical projection layer of 10 km grid cells (100 km^2), resulting in 3866 grid cells for analysis. A 10 km grid cell size was chosen after preliminary analyses of the total dataset revealed a mean distribution of 1 survey per 100 km^2. All spatial data processing was done using ArcGIS version 9.3 (ESRI Inc. 2008).

(B)Volunteer survey data

We used a subset of bird surveys from 1998 to 2011 obtained from the New Atlas of Australian Birds, an on-going project of which approximately 95% of the surveys are collected by volunteers throughout Australia (Barrett *et al.* 2003). Records were checked for reliability and surveys with no recorded coordinate system or less than 5-km locational accuracy were discarded. The dataset was split into three time periods: a) 1998–2002, the main atlas period during which data were collected for publication of species range maps, with an emphasis on covering as many different sites as possible (Barrett *et al.* 2003), b) 2003–2007, the bulk of surveys conducted after the atlas was published when volunteers were allowed to visit any site of their choosing (hereafter termed ‘post-atlas’ surveys), comprising the validation data, and c) 2008–2011, the test data for ‘the future’. The volunteer survey dataset was overlaid with a 10 km grid of the study area to calculate, for each grid cell in each time period: a) the number of surveys per year, b) the identity of species detected each year and the total number of species detected per year, and c) the identity and number of threatened species (listed as threatened fauna on the Wildlife Conservation Act 1950 in WA) detected per year.

(B)Data analysis

All statistical analyses were carried out in R version 2.11.1 (R Development Core Team 2010).

(C)Predicting future survey distribution from past surveys

Our first aim was to determine if the spatial distribution of past surveys can predict the distribution of future surveys. To do this, we related the spatial distribution of survey effort (number of surveys per grid cell) from the post-atlas period (2003–2007; response variable) to the distribution in the main atlas collection period (1998–2002; explanatory variable). We used generalised linear modelling (GLM) with a Poisson distribution to test for the significance of the relationship and thus the ability of the distribution of past surveys to predict the future locations of volunteer surveys. The residuals of all models were tested to ensure no model assumptions were violated, including tests for over-dispersion and Moran's I test for spatial autocorrelation.

(C)Volunteer motivations

Our second aim was to determine factors that motivate volunteers to survey different parts of the landscape, using a species distribution modelling approach (Guisan & Zimmermann 2000; Elith & Leathwick 2009). Human behaviour is complex, and to fully understand the reasons for volunteers to survey an area we would ideally have actual questionnaire data that the volunteers have provided, however these data have not been collected on a broad scale and are difficult to collect after the event. We therefore used coarse-scale landscape surrogates to represent the factors that motivate volunteers to survey in different areas. We developed a number of hypotheses describing potential motivations for volunteers to survey different areas based on previous literature on biases in citizen science datasets, resulting in 20 models in total (see Table 1). Generalised linear modelling with a Bernoulli distribution and a logit link was used to fit each model to the survey occurrence data for each grid cell in 2003–2007. The residuals of all models were again tested to ensure no model assumptions were violated, including tests for over-dispersion and spatial autocorrelation. For explanatory variables, we used 20 environmental factors describing landscape characteristics, and one variable describing the richness of threatened birds detected in the main atlas period (Table 2), as previous studies indicate that many volunteers seek out threatened bird species (Booth *et al.* 2011; Tulloch & Szabo accepted). Preliminary analyses found strong correlations between the potential explanatory variables of distance from Perth, distance from towns, and road density (Pearson's product-moment correlation, $r^2 > 0.7$, $P < 0.001$), so models containing these variables were simplified into separate models to avoid colinearity of variables (see Table 1). In order to allow direct comparison between explanatory variables measured in different units, continuous explanatory variables were first standardised by subtracting the mean value and dividing by two standard deviations (Gelman & Hill 2007). Hypotheses were compared in an information-theoretic framework using AIC model selection (Burnham & Anderson 2002), and the best-supported model was used to predict the probability of survey occurrence for every grid cell in the region.

We validated the predictive performance of the best-supported model with a random selection of 10% of the survey data from the 2003–2007 dataset held back from analyses. We explored the agreement between model predictions and observations using calibration diagrams (Pearce & Ferrier 2000), and the distribution of predicted values for surveyed and unsurveyed cells (Elith & Leathwick 2009). The Area Under the Receiver Operator Curve (AUC) – as a measure of rank-correlation– was calculated to evaluate the quality of the predictions. A high AUC value indicates that high predicted scores tend to be areas of known presence and lower model prediction scores tend to be areas in which the surveyors are known to be absent (or a random point). An AUC score of 0.5 means that the model is as good as a random guess. We also explored model refinement, which relates to the total range of predictions produced by the model. A model is well refined if predictions cover the full probability range, with predicted values near both one and zero (Pearce & Ferrier 2000).

(C)Predicting survey distribution based on volunteer motivations

Our third aim was to use our findings on volunteer motivations to predict the probability of areas being surveyed later. To address this aim, the best-supported model of volunteer motivations was used to predict the probability of areas being surveyed in 2008–2011, and predictive performance was evaluated as above. We compared our predicted values for our best model for 2003–2007 data with a new model that has the same explanatory variables but new data (2008–2011 surveys), to see if the same cells were predicted to be surveyed. This resulted in a new probability of survey (A_i = Actual P_i) that can be compared with the old (predicted P_i), where P_i represented the chance a cell was surveyed in the designated time period as informed by the model. If our model predicts well, we should see a strong one-to-one relationship between A_i and P_i, and we have higher confidence in predicting survey effort in later years.

(C) Prioritisation of future surveying

Our final aim was to develop a protocol to prioritise future surveying effort which accounts for the probability of an area being surveyed by volunteers. In order to do this we first envisaged three possible surveying goals:

1) Equal representation
   Target: At least two surveys per grid cell (since previous studies (Cunningham *et al.* 1999; Field *et al.* 2005) show that more than one survey is needed to increase the likelihood of detecting all species).
2) Stratified habitat representation
   Target: At least one survey per habitat per grid cell (Beard 1980a; b), so that the number of surveys per grid cell ≥ number of habitat types per grid cell.
3) Stratified protected area representation
   Target: Surveys per grid cell ≥ number of protected areas per grid cell (WA Department of Environment and Conservation).

By applying these goals to the landscape we explored the potential benefits of investing in professional sampling in survey gaps. We first calculated the number of surveys already achieved towards these goals (i.e. current status of each grid cell) by overlaying the 1998–2002 and 2003–2007 bird survey datasets on the relevant spatial layers for each goal. The number of surveys s_a^g required in a spatial unit a to achieve each surveying goal g is the difference between the number sought after by the goal (see above definitions) and the number currently achieved, with a value of zero for a spatial unit indicating that unit has achieved the goal (and negative values therefore converted to zero). We used the predicted probability of volunteer surveys for 2008–2011 to determine the probability of a survey not taking place (1 – Probability of being surveyed by a volunteer) for each grid cell. We were then able to derive benefit functions for investing in additional surveying to meet each of our goals (1–3) based on the probability of a volunteer survey not taking place, in which surveys unlikely to be conducted by volunteers, but which will contribute to our goal, will have a high benefit:

$$\text{Survey benefit} = s_a^g * \Pr(\text{no volunteer surveys}).$$

Finally, we calculated the cost of meeting a surveying goal with and without accounting for the probability of cells not being surveyed by volunteers. We calculated the cost of surveying using the third scenario as an example (stratified protected area representation). The cost per sampling unit was calculated using the equation:

$$\text{Cost per sampling unit} = \sum x_a.c_a.s_a$$

Where a is the target area (here a protected area) in which sampling is required to achieve a goal, x_a is the action of selecting or not selecting target a for survey (x_a is 0 or 1), c_a is the cost for one 20 minute 2-ha standardised survey in target a, and s_a is the number of surveys required in each target a. For this study c_a was set at AU$50 across all habitats and s was 1. To determine the initial cost in each sampling unit without taking into account the future efforts of volunteers, x_a was set to 1 for all protected areas still unsurveyed, and 0 for all protected areas already surveyed. We targeted protected areas with a 70% chance of not being surveyed. The cost per grid cell was then re-calculated to prioritise only the protected areas with at least 70% probability of no volunteer surveys. We were therefore able to calculate the cost-saving that would result from incorporating knowledge of expected volunteer behaviour to prioritise future surveying.

(A)RESULTS

Our first aim was to explore the relationship between the spatial distribution of surveys from the main atlas period (1998–2002) and the post-atlas period (2003–2007) to assess the ability of past survey distribution to predict later distribution. A GLM showed that the distribution of the number of surveys per grid cell during the post-atlas period in 2003–2007 (response variable) is positively associated with the number of surveys recorded during the main atlas period in 1998–2002 (explanatory variable; deviance explained 34.85%, $\beta = 0.80$, SE = 0.02; Fig. 1, Table S2).

Our second aim was to test hypotheses describing volunteer motivations for surveying in different areas. We found that the best-supported model for survey occurrence in 2003–2007 was the model that described “conservation concern” motivation, with an AIC weight of 1 ranking it conclusively above all other models (Table 3, Figs. 2 and S1(a)). This model accounted for 13.09% of the deviance (Fig. 2(a, b, c); see Table 2 for description of explanatory variables). The probability of survey occurrence in a grid cell increases with the number of protected areas being managed for conservation (“protected areas”: $\beta = 0.83$, SE = 0.08;

Table 4, Fig. S1(b)), the number of habitat types in that grid cell ("habitat diversity": $\beta = 0.54$, SE = 0.08; Table 4, Fig. S1(c)) and when at least one threatened species was recorded during the main atlas period 1998–2002 ("threatened species presence": $\beta = 1.24$, SE = 0.08; Table 4, Fig. S1(d)). The mean number of protected areas per grid cell is 2.65 ± 0.14 SE (with a probability of survey of 0.22), with the predicted probability of a survey in a grid cell only 0.10 if there is less than one protected area per cell, but increasing to 0.47 when there are more than 100 protected areas (Fig. 2(a)). The mean number of habitat types per grid cell is 3.90 ± 0.03 SE (probability of survey of 0.22), and the predicted probability of a volunteer surveying is only 0.09 if there is only one habitat type per cell, increasing to 0.39 when there are more than 13 habitats (Fig. 2(b)). The probability of a volunteer survey occurring if a threatened species has previously been recorded in a grid cell is 0.37, but if no threatened species has been recorded there the probability is reduced to 0.15 (Fig.2(c)).

The predicted values for cells at which surveys were recorded in 2003–2007 were, on average, higher than those for unsurveyed cells, indicating a good discrimination capability of the best model (Fig. S2). This was confirmed by a plot of the Receiver Operating Curve, with an AUC of 0.83 (Fig. S2). The refinement of the values predicted by the model was also good, with predictions ranging from 0.06 to 0.88.

Our third aim was to test the best-supported model (threatened species and protected areas motivation) using recently-collected (since 2008) atlas data. The discrimination ability of the model was assessed visually by comparing the distribution of predicted probabilities for occupied cells with the distribution of the predicted probabilities for unoccupied cells (Fig. S3(a)). Predicted values for cells at which the surveys were recorded between 2008 and 2011 were, again, higher than those for unsurveyed cells (Fig. S3(b)), which was confirmed by an AUC of 0.73 (Fig. S3(c)). There were strong linear relationships between A_i (applying the best-supported model to data from 2008–2011) and predicted P_i (using data from 2003–2007) for cells not visited yet ($r^2 = 0.995$, $P < 0.001$) and cells visited in 2008–2011 ($r^2 = 0.97$, $P < 0.001$) (Fig. 3).

Our final aim was to apply different monitoring goals to an expected benefits function for undertaking professional surveys, given the distribution of volunteer efforts. We found different spatial distributions of effort allocation according to the goal (Fig. 4). To achieve a goal of equal representation across grid cells (two surveys per grid cell), benefits were distributed across the inland parts of the study area (Fig. 4(a)); these cells had lower survey counts in 2003–2007 (Fig. 1(b)). The expected benefit increased linearly with the probability of a cell not being surveyed ($r^2 = 1$, $P < 0.001$; Fig. 5(a)), and the mean benefit was 1.51 ± 0.34 SD. To achieve habitat stratification goals for monitoring (one survey per habitat type in each grid cell), expected monitoring benefits were scattered across the landscape but with generally low values around Perth and surrounds (Fig. 4(b)). There was a slight positive linear relationship between the expected benefit and the probability of a survey not occurring ($r^2 = 0.02$, $P < 0.001$; Fig. 5(b)), with a mean benefit of 2.81 ± 1.21 SD. When surveys were prioritised in protected areas (one survey per protected area in each grid cell), the highest benefits were located in the south-western corner of the study area where there is a high density of small fragmented protected areas (Fig. 4(c)), and there was a slight positive linear relationship between benefits and probability of a survey not taking place ($r^2 = 0.02$, $P < 0.001$; Fig. 5(c)).

The mean benefit was 1.91 ± 4.09 SD. When converted to selection frequencies of cells over the mean value for each benefit function, there were only 131 cells (3%) selected by all three benefit functions, with the majority selected by either one (2006 cells, 52%) or two (1307 cells, 34%) functions; 422 cells (11%) were not selected by any benefit function.

The costs to achieve surveying goal 3 (one survey in each protected area in each 100km^2 grid cell) in each grid cell varied depending on the number of protected areas already surveyed. Protected areas cover approximately 68,000 km^2 (19% of the study area) and contain 18.5% of records. This equates to 2313 grid cells containing protected areas, of which only 163 (7%) satisfied the target of one survey per protected area per cell. At least 85% of protected areas had never recorded a 2-ha survey, despite 20 of these being designated National Parks. The total budget for surveying all protected areas not yet surveyed in every grid cell was AU$643,850. Surveying in only the protected areas with more than 70% chance of not being surveyed by volunteers reduced the estimated total cost by almost 75% to AU$164,750.

(A)DISCUSSION

To evaluate and manage current levels of biodiversity loss we need to quantify biodiversity composition at large spatial and temporal scales (Henry *et al.* 2008). However, biological surveys are expensive and time consuming. The cost of managing and monitoring Australia's 155 threatened bird taxa alone is estimated at least US$10.2 million annually, and available government funding is less than a third of this amount (Garnett *et al.* 2003). Although many papers discuss optimal sampling of the environment, few studies have focused on the behaviour of the people doing the monitoring. This is particularly important for large-scale datasets such as atlases, which accumulate vast quantities of data through the efforts of volunteers. This paper has shown how we can apply the concept of 'species distribution modelling' to humans, by developing models of volunteer motivations to optimally invest in additional surveys or modify human behaviour to achieve specific outcomes. By looking at past motivations for volunteer bird surveyors, we can predict their future behaviour and use these predictions to determine the areas that are unlikely to be visited by volunteers. We have demonstrated that defining a clear objective for monitoring is crucial as the sites selected for future investment in surveying can vary greatly depending on the specified goal (Figs. 4 and S1).

Broad-scale surveys (e.g. atlases) and monitoring programs (e.g. breeding bird surveys), are used increasingly to document declines in populations worldwide as well as to evaluate the efficacy and efficiency of management actions and policies (Gates & Donald 2000; Yoccoz *et al.* 2001; Telfer *et al.* 2002; Kotze & O'Hara 2003). However, reviews of existing monitoring programs have highlighted a lack of well-articulated objectives (Dale & Beyeler 2001; Legg & Nagy 2006; Lindenmayer & Likens 2009; 2010), which can lead to the wrong variables being measured in the wrong place at the wrong time, and result in datasets that do not have sufficient statistical power to answer important questions. Driving volunteer surveying or monitoring programs by well-formulated and tractable objectives can help to avoid inefficient and ineffective actions and wasting of limited resources (Salzer & Salafsky 2006; Grantham *et al.* 2009; Lindenmayer & Likens 2009). We devised three goals for surveying the birds of the south-western Australian biodiversity hotspot, which resulted in different spatial patterns of the benefits of investing in future surveys (Fig. 4). The different goals demand different management interventions. For example, ensuring that all environmental space has been covered might be important for developing accurate species distribution models that can be used to prioritise conservation planning decisions (Loiselle *et al.* 2003; Carvalho *et al.* 2010). In contrast, identifying protected areas that are unlikely to be visited due to their location or lack of publicly-available information can assist with evaluating protected area performance. In a social science context this might involve understanding the level of public knowledge of protected areas to evaluate their potential for leveraging support for conservation activities or impacting policy decisions.

Knowing where people voluntarily undertake surveys helps us devise cost-effective strategies for a particular objective. For example, to achieve a goal of stratified protected area representation in our study landscape (Fig. 4(c)), there was a 75% reduction in the cost of the planned survey program if only protected areas predicted to have a 70% chance of not being surveyed by volunteers were selected for professional surveys. These results show that by setting clear objectives, and accounting for information on the benefits and costs of an action, such as monitoring of birds by volunteers, organizations wishing to implement potentially expensive surveying programs across large scales can improve the effectiveness of future efforts by investing limited funding where it is most needed.

Understanding the motivations and reasons behind volunteer behaviour is crucial to enable better management of the volunteers and data generated by them. Given the complex nature of human behaviour, ideally data on motivation, collected from volunteers, would be used to investigate this. However, at present these data are not usually available at the landscape scale. An alternative methodology, as used here, is to use coarse environmental variables as correlates of volunteer distribution and infer from this the motivations for visiting particular locations. It is important to bear in mind, when interpreting our results, that the inferences are based on correlational and not mechanistic models. The benefit of this approach is that it can be applied to any atlas or other large scale survey data-set now without further data collection, and the results can be used to prioritise further survey or monitoring investment. The use of species distribution modelling with an AIC model selection framework is ideal to test these ideas with this type of data because we are able to compare multiple hypotheses and highlight the model/s best supported by the data, and where relevant, make inference across multiple models to reduce model selection uncertainty. However, it is important to stress that the ideal would be to have demographic data, and explicit data on volunteer motivations in order to better elucidate why volunteers go where they go, and in particular whether they could be incentivised to go elsewhere. For example, birdwatchers are not acting independently of one another, and indeed may travel together, visit sites recommended by friends or engage in competitive bird watching, all of which will affect the distribution of surveys in the landscape. Dependence between volunteers (e.g. travelling together) will also in some circumstances increase the chance of achieving multiple surveys per spatial unit and would be useful to explore if the data were available. Although this information is not collected by the dataset we used, some citizen science projects already collect data on dependencies through the ability of online users to ‘share’ data with members from their survey group (e.g. eBird), which will allow more advanced kinds of modelling that incorporate volunteer dependencies to be explored. We suggest that these types of data could be routinely collected as part of citizen science projects, such as by the inclusion of a short questionnaire within the data forms that volunteers are already completing to record their data, and that this should be done as a matter of priority in order to inform future survey and monitoring efforts.

That said, our analyses have highlighted that the best model of volunteer behaviour in this study described motivations to visit areas or species of conservation concern– i.e. that threatened species, habitat diversity and protected areas are the major drivers for bird surveying. This corresponds with a previous study in Australia (Weston *et al.* 2006), which found strong conservation motivations in a survey of volunteers for the Threatened Bird Network. Similarly, research in other parts of the world shows volunteers are attracted to areas of high diversity (Parnell *et al.* 2003; Romo *et al.* 2006) and also to rare species (Booth *et al.* 2011). The importance of previous sightings of threatened species as a motivating variable in our model is supported by previous research that indicates that volunteers often return to the same location where they previously saw a threatened species (Booth *et al.* 2011; Tulloch & Szabo accepted), or share their knowledge of where they know a threatened or rare species is located (e.g. Australia: http://birding-aus.org/; South Africa: http://groups.yahoo.com/group/sabirdnet; United Kingdom: http://www.birdnetinformation.co.uk). The importance of protected areas for volunteer motivation is also not surprising. In this study, grid cells with high numbers of protected areas were generally more visited than those with few (e.g. one large protected area). This might be due to accessibility issues such as road access, which has been shown to bias the distribution of volunteer surveys (Longmore 1986; Reddy & Dávalos 2003; Romo *et al.* 2006; Botts *et al.* 2010). Interestingly, the models including accessibility performed substantially worse than our “conservation concern” model (Tables 3 and S2), highlighting that focusing solely on accessibility as an explanation for the distribution of volunteers in the landscape may result in an incomplete understanding of the true motivations..

The models of motivations for human behaviour that we developed in this study were contextualised through our knowledge of the study area and the bird surveyors in Australia and in other parts of the world (see Table 1). There is a large body of literature that emphasises the need to include ecological reasoning when choosing appropriate models. Different scenarios, different availability of environmental information or different landscapes might result in a different selection of explanatory variables for the same hypotheses in another region or for another species of interest. The datasets we have used in this study are generally freely available, and therefore easily accessible to both government and non-government organisations interested in prioritising actions in a cost-effective way. The models we have developed for our region can therefore be tested on other areas with similar immediate needs of filling knowledge gaps, e.g. for conservation planning. Other volunteer-collected databases for both birds and other species, in particular in the southern hemisphere and developing regions where the areas to be covered are large, have shown data incompleteness and/or strong biases in the distribution of surveys, including in the Amazon (Williams *et al.* 1996), Australia (Longmore 1986; Margules *et al.* 1994; Szabo *et al.* 2007), and parts of Africa (Funk & Richardson 2002; Reddy & Dávalos 2003; Botts *et al.* 2010). These regions have a relatively short history of scientific research as well as lower population densities compared to Europe and the USA, resulting in fewer people potentially available for environmental volunteering (Gibbons *et al.* 2007; Dunn & Weston 2008). For example, in the USA, the New York State Breeding Bird Atlas (Hampe 2004) mobilised over 1200 volunteers, which represents 151 volunteers per 100 km^2, more than 100 times the number available for the South African Frog Atlas (420 volunteers, 0.014 per 100 km^2) and South African Bird Atlas Project 2 (1029 volunteers, 0.0343 per 100 km^2), both of which have biased datasets with gaps in knowledge (Robertson & Barker 2006; Botts *et al.* 2010). However, these under-sampled regions have an equal if not greater need for standardised and representative survey datasets for conservation planning and assessment, especially for taxa other than birds that may not have such a high general appeal. Even with clear directions and communication from the data custodians it would be difficult for volunteers to cover all of the 3 million km^2 of South Africa without some form of survey prioritisation. Decisions on the hypotheses to be tested for the prioritisation approach that we have developed, should be informed by knowledge of the region and its volunteers, and the choice of explanatory variables based on the best available information as well as knowledge of the study species (e.g. frog atlases will most likely require variables the describe the habitat and/or climate limitations of amphibians). In cases of low data availability, an alternative approach such as Bayesian Belief Networks informed through expert elicitation, might be more appropriate (Kuhnert *et al.* 2010).

Systematic sampling in survey gaps is required to obtain more reliable volunteer-collected datasets that are less spatially biased (Balmford *et al.* 2003; Balmford *et al.* 2005). We have demonstrated one way of filling these gaps through professional surveys. Additional approaches to fill these gaps could include using incentives and marketing to persuade future volunteers to travel away from their preferred areas (e.g. roads, urban areas, ‘interesting’ habitats). Providing access to the accumulated data and some of its interpretation can be an important motivating tool to achieve this. A first important step for atlas coordinators is to publish maps for survey distribution to volunteers that highlight grid cells (or other target areas) with no records, or to update record maps (and data needs) regularly online similarly to the South African Bird Atlas Project 2 (http://sabap2.adu.org.za/) and eBird (http://ebird.org/content/ebird/). Some databases (e.g. eBird) have had some success encouraging volunteers to monitor unsurveyed areas using ‘patch competitions’, where volunteers compete to find the most species in their own chosen or allocated area (Sullivan *et al.* 2009). However, it is hard to get people to survey in places that they have no interest in visiting. In these areas, if data are urgently required (e.g. for conservation planning needs or environmental impact assessment), the only option available might be to pay for surveys. This is exactly what was done for the New Catalan Breeding Bird Atlas (Hampe 2004), albeit over a much smaller area of only 32,000 km^2 (a tenth of our study area). In much larger areas (e.g. Australia), it will not be feasible to fill all the gaps with expensive professional surveys, and a prioritisation approach as shown here could be appropriate.

There are considerable benefits to professional surveying, namely the ability to standardize and direct surveys to the areas where they are most needed. We present a method to prioritise additional organised surveys, in which areas remaining unsurveyed are prioritised for professional surveying based on the particular needs and goals of data collection.

Uneven distribution of surveys can lead to more than just gaps in space. Recent findings show that the representation of bird species in the current Atlas of Australian Birds is not even, rather it is biased towards species found in urban areas and preferred habitats (Szabo *et al.* 2007; Tulloch & Szabo accepted), resulting in fewer species being detected in some areas than expected. Increased sample size and a more representative spatial distribution of sampling units in surveying and monitoring programs (including atlases) can ensure that we obtain information on all species, not just those located in the areas volunteers are most interested in visiting. This should lead to more powerful predictive species distribution models (Brotons *et al.* 2007), more accurate species richness maps for conservation assessments, policy and planning, and more efficient use of limited resources for management and monitoring.

Ultimately we have developed a basic species distribution model (SDM) that assumes that humans behave similarly to each other. We do not deny that human behaviour is too complex to analyse with simple models, but given that we do not yet have this kind of information for all of the volunteers, this is a useful 'first pass'. Uncertainty pervades in all SDMs: in the models we choose, the parameters we select, and the way we interpret them (Elith *et al.* 2002; Barry & Elith 2006; Pearson *et al.* 2006; Elith & Leathwick 2009). For example, some volunteers in our models most likely are not acting independently of each other. Due to the coarse scale of our analysis we do not account for this, similarly it is difficult to account for biotic interactions in species distribution modelling (Hampe 2004; Guisan *et al.* 2006). Increases in technology and crowd sourcing mean that more detailed data on the volunteers themselves should become easier to collect in the future. Planners of citizen science programs such as atlases should ensure that they harness the advances in technology that allow them to collect vital demographic data, as well as information on motivations of volunteers, which will assist with understanding volunteer motivations and ultimately planning investment in to monitoring programs in to the future.

(A)CONCLUSION

Immediate action to address key gaps in volunteer-collected databases such as atlases will ensure that they are more powerful tools for decision-making. It is clear that the actions of volunteer bird surveyors in conjunction with organisations that fund targeted monitoring programs will play a major role in closing knowledge gaps and enabling continued monitoring of key habitats, species, and their threats. This is the first study to predict the probability of volunteer surveys occurring across a landscape, and use these probabilities to develop benefit functions for prioritising surveying efforts that explicitly account for the objective as well as the likelihood of a volunteer survey occurring there. The approach we describe uses a simple and repeatable species distribution modelling technique that can be carried out with any large-scale datasets in which volunteers choose where to survey. We have also highlighted the need to routinely collect some demographic and motivational data as part of such datasets. We have shown that incorporating knowledge of the motivations of volunteers can assist with prioritising surveying and on-going monitoring efforts, by reducing the number of planned survey sites and their associated survey costs. Prioritising resource allocation for monitoring and directing more money to management of threatened species will deliver better conservation outcomes.

Acknowledgements

This research was conducted with the support of funding from the Australia Research Council (ARC) Centre of Excellence for Environmental Decisions and an ARC Linkage Grant. A. Tulloch was supported during her research by an Australian Postgraduate Award, a Gondwana Link scholarship, and by a 2011 Birds Australia Stuart Leslie Bird Research Award. We are grateful to all the volunteers that contributed to the Atlas of Australian Birds, and to Birds Australia and the Western Australian Department of Environment and Conservation for the use of their data. We thank four anonymous reviewers for their constructive comments.

Biosketch

Ayesha Tulloch is a PhD candidate with the Environmental Decisions Group at the University of Queensland, and the Centre of Excellence for Environmental Decisions (CEED, http://ceed.edu.au/). Her PhD research focuses on integrating disciplinary perspectives (economic, social, and environmental) to evaluate approaches for prioritising conservation investments in multiple stakeholder landscapes. Her research interests include optimal monitoring and management of multiple threats and species, with a particular interest in invasive predators, socio-ecological systems analysis, and decision-making processes for conservation resource allocation.

LITERATURE CITED

Balmford, A., Crane, P., Dobson, A., Green, R. E. & Mace, G. M. (2005). The 2010 challenge: Data availability, information needs and extraterrestrial insights. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **360**, 221-228.

Balmford, A., Green, R. E. & Jenkins, M. (2003). Measuring the changing state of nature. *Trends in Ecology & Evolution*, **18**, 326-330.

Barrett, G. (2000). Birds on farms: ecological management for agricultural sustainability. *Wingspan Supplement*, **10**, 1-16.

Barrett, G. W., Silcocks, A., Barry, S., Cunningham, R. & Poulter, R. (2003). *The New Atlas of Australian Birds,* Melbourne, Birds Australia.

Barry, S. & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413-423.

Bas, Y., Devictor, V., Moussus, J.-P. & Jiguet, F. (2008). Accounting for weather and time-of-day parameters when analysing count data from monitoring programs. *Biodiversity and Conservation*, **17**, 3403-3416.

Battersby, J. E. & Greenwood, J. J. D. (2004). Monitoring terrestrial mammals in the UK: past, present and future, using lessons from the bird world. *Mammal Review*, **34**, 3-29.

Beard, J. S. (1980a). A New Phytogeographic Map of Western Australia. *Research notes of the Western Australian Herbarium*, **3**, 37-58.

Beard, J. S. (1980b). *Vegetation survey of Western Australia. Western Australia 1:1 000 000 vegetation series. Swan Sheet 7,* Nedlands, W.A. , University of Western Australia Press, with support of the Interim Council for the Australian Biological Resources Study.

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K. & Mace, G. M. (2010). Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data. *PLoS Biology*, **8**, e1000385.

Booth, J. E., Gaston, K. J., Evans, K. L. & Armsworth, P. R. (2011). The value of species rarity in biodiversity recreation: A birdwatching example. *Biological Conservation*, **144**, 2728-2732.

Botts, E. A., Erasmus, B. F. N. & Alexander, G. J. (2010). Geographic sampling bias in the South African Frog Atlas Project: implications for conservation planning. *Biodiversity and Conservation*, **20**, 119-139.

Brotons, L., Herrando, S. & Pla, M. (2007). Updating bird species distribution at large spatial scales: applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions*, **13**, 276-288.

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach,* New York, Springer-Verlag.

Carlson, M. & Schmiegelow, F. (2002). Cost-effective sampling design applied to large-scale monitoring of boreal birds. *Conservation Ecology*, **6**.

Carvalho, S. B., Brito, J. C., Pressey, R. L., Crespo, E. & Possingham, H. P. (2010). Simulating the effects of using different types of species distribution data in reserve selection. *Biological Conservation*, **143**, 426-438.

Cunningham, R. B., Lindenmayer, D. B., Nix, H. A. & Lindenmayer, B. D. (1999). Quantifying observer heterogeneity in bird counts. *Australian Journal of Ecology*, **24**, 270-277.

Dale, V. H. & Beyeler, S. C. (2001). Challenges in the development and use of ecological indicators. *Ecological Indicators*, **1**, 3-10.

Danielsen, F., Burgess, N. D., Balmford, A., Donald, P. F., Funder, M., Jones, J. P. G., Alviola, P., Balete, D. S., Blomley, T., Brashares, J., Child, B., Enghoff, M., Fjeldsa, J., Holt, S., Hubertz, H., Jensen, A. E., Jensen, P. M., Massao, J., Mendoza, M. M., Ngaga, Y., Poulsen, M. K., Rueda, R., Sam, M., Skielboe, T., Stuart-Hill, G., Topp-Jorgensen, E. & Yonten, D. (2009). Local Participation in Natural Resource Monitoring: a Characterization of Approaches. *Conservation Biology*, **23**, 31-42.

Department of Agriculture and Food (2006). Land use in Western Australia. Spatial data layer - ESRI ArcInfo surface.

Donald, P. F. & Fuller, R. J. (1998). Ornithological atlas data: a review of uses and limitations. *Bird Study*, **45**, 129-145.

Dunn, A. M. & Weston, M. A. (2008). A review of terrestrial bird atlases of the world and their application. *Emu*, **108**, 42-67.

Elith, J., Burgman, M. A. & Regan, H. M. (2002). Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, **157**, 313-329.
Elith, J. & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology Evolution and Systematics*, **40**, 677-697.
ESRI Inc. (2008). *ArcGIS 9.3,* Redlands, CA, Environmental Systems Research Institute Inc.
Etterson, M. A., Niemi, G. J. & Danz, N. P. (2009). Estimating the effects of detection heterogeneity and overdispersion on trends estimated from avian point counts. *Ecological Applications*, **19**, 2049-2066.
Field, S. A., Tyre, A. J. & Possingham, H. P. (2005). Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management*, **69**, 473-482.
Fleishman, E. & Murphy, D. D. (2009). A Realistic Assessment of the Indicator Potential of Butterflies and Other Charismatic Taxonomic Groups. *Conservation Biology*, **23**, 1109-1116.
Funk, V. A. & Richardson, K. S. (2002). Systematic Data in Biodiversity Studies: Use It or Lose It. *Systematic Biology*, **51**, 303-316.
Garnett, S., Crowley, G. & Balmford, A. (2003). The Costs and Effectiveness of Funding the Conservation of Australian Threatened Birds. *Bioscience*, **53**, 658-665.
Garnett, S., Szabo, J. K. & Dutton, G. (2011). *The Action Plan for Australian Birds 2010,* Collingwood, CSIRO Publishing.
Gates, S. & Donald, P. F. (2000). Local Extinction of British Farmland Birds and the Prediction of Further Loss. *Journal of Applied Ecology*, **37**, 806-820.
Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Heirarchical Models,* Cambridge, Cambridge University Press.
Gerber, L. R., Beger, M., McCarthy, M. A. & Possingham, H. P. (2005). A theory for optimal monitoring of marine reserves. *Ecology Letters*, **8**, 829-837.
Gibbons, D. W., Donald, P. F., Bauer, H.-G., Fornasari, L. & Dawson, I. K. (2007). Mapping avian distributions: the evolution of bird atlases. *Bird Study*, **54**, 324 - 334.
Grantham, H. S., Wilson, K. A., Moilanen, A., Rebelo, T. & Possingham, H. P. (2009). Delaying conservation actions for improved knowledge: how long should we wait? *Ecology Letters*, **12**, 293-301.
Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R. & Hastie, T. (2006). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43**, 386-392.
Guisan, A. & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.
Hampe, A. (2004). Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, **13**, 469-471.
Hauser, C. E., Pople, A. R. & Possingham, H. P. (2006). Should managed populations be monitored every year? *Ecological Applications*, **16**, 807-819.
Henry, P. Y., Lengyel, S., Nowicki, P., Julliard, R., Clobert, J., Celik, T., Gruber, B., Schmeller, D., Babij, V. & Henle, K. (2008). Integrating ongoing biodiversity monitoring: potential benefits and methods. *Biodiversity and Conservation*, **17**, 3357-3382.
Hopper, S. D. & Gioia, P. (2004). The Southwest Australian Floristic Region: Evolution and conservation of a global hot spot of biodiversity. *Annual Review of Ecology Evolution and Systematics*, **35**, 623-650.
Jones, J. P. G. (2011). Monitoring species abundance and distribution at the landscape scale. *Journal of Applied Ecology*, **48**, 9-13.
Joseph, L. N., Field, S. A., Wilcox, C. & Possingham, H. P. (2006). Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology*, **20**, 1679-1687.
Kery, M., Spillmann, J. H., Truong, C. & Holderegger, R. (2006). How biased are estimates of extinction probability in revisitation studies? *Journal of Ecology*, **94**, 980-986.
Kotze, D. J. & O'Hara, R. B. (2003). Species Decline: But Why? Explanations of Carabid Beetle (Coleoptera, Carabidae) Declines in Europe. *Oecologia*, **135**, 138-148.
Kuhnert, P. M., Martin, T. G. & Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, **13**, 900-914.
Küper, W., Sommer, J. H., Lovett, J. C. & Barthlott, W. (2006). Deficiency in African plant distribution data – missing pieces of the puzzle. *Botanical Journal of the Linnean Society*, **150**, 355-368.

Legg, C. J. & Nagy, L. (2006). Why most conservation monitoring is, but need not be, a waste of time. *Journal of Environmental Management*, **78**, 194-199.

Lindenmayer, D. B. & Likens, G. E. (2009). Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends in Ecology & Evolution*, **24**, 482-486.

Lindenmayer, D. B. & Likens, G. E. (2010). *Effective ecological monitoring,* Collingwood, VIC, CSIRO Publishing.

Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G. & Williams, P. H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, **17**, 1591-1600.

Longmore, R. (1986). Atlas of elapid snakes of Australia). Australian Government Publishing Service, Canberra.

Loyola, R. D., Oliveira-Santos, L. G. R., Almeida-Neto, M., Nogueira, D. M., Kubota, U., Diniz, J. A. F. & Lewinsohn, T. M. (2009). Integrating Economic Costs and Biological Traits into Global Conservation Priorities for Carnivores. *PLoS ONE*, **4**, 9.

Margules, C. R., Austin, M. P., Mollison, D. & Smith, F. (1994). Biological Models for Monitoring Species Decline: The Construction and Use of Data Bases. *Philosophical Transactions: Biological Sciences*, **344**, 69-75.

Mittermeier, R. A., Myers, N., Thomsen, J. B., da Fonseca, G. A. B. & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: Approaches to setting conservation priorities. *Conservation Biology*, **12**, 516-520.

Moilanen, A., Wilson, K. & Possingham, H. P. (2009). *Spatial Conservation Prioritization: Quantitative Methods and Computational Tools,* New York, Oxford University Press.

Munson, M. A., Caruana, R., Fink, D., Hochachka, W. M., Iliff, M., Rosenberg, K. V., Sheldon, D., Sullivan, B. L., Wood, C. & Kelling, S. (2010). A method for measuring the relative information content of data from different monitoring protocols. *Methods in Ecology and Evolution*, **1**, 263–273.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853-858.

Osborne, P. E. & Tigar, B. J. (1992). Interpreting bird atlas data using logistic models: an example from Lesotho, southern Africa. *Journal of Applied Ecology*, **29**, 55-62.

Parnell, J. A. N., Simpson, D. A., Moat, J., Kirkup, D. W., Chantaranothai, P., Boyce, P. C., Bygrave, P., Dransfield, S., Jebb, M. H. P., Macklin, J., Meade, C., Middleton, D. J., Muasya, A. M., Prajaksood, A., Pendry, C. A., Pooma, R., Suddee, S. & Wilkin, P. (2003). Plant collecting spread and densities: their potential impact on biogeographical studies in Thailand. *Journal of Biogeography*, **30**, 193-209.

Parsons, H. & Major, R. E. (2004). Bird interactions in Sydney gardens: some initial findings of the Birds in Backyards program. *Urban Wildlife: More Than Meets the Eye*. (ed. by D. Lunney and S. Burgin), pp. 211-215. Royal Zoological Society of NSW, Mosman, NSW.

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

Pearson, R. G., Thuiller, W., Araujo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P. & Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704-1711.

Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181-197.

Pomeroy, D., Tushabe, H. & Cowser, R. (2008). Bird atlases - how useful are they for conservation? *Bird Conservation International*, **18**, S211-S222.

R Development Core Team (2010). *R: A language and environment for statistical computing, reference index version 2.11.1,* Vienna, Austria. URL http://www.r-project.org, R Foundation for Statistical Computing.

Rachlow, J. L. & Svancara, L. K. (2006). Prioritizing habitat for surveys of an uncommon mammal: A modeling approach applied to pygmy rabbits. *Journal of Mammalogy*, **87**, 827-833.

Reddy, S. & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719-1727.

Regan, H. M., Hierl, L. A., Franklin, J., Deutschman, D. H., Schmalbach, H. L., Winchell, C. S. & Johnson, B. S. (2008). Species prioritization for monitoring and management in regional multiple species conservation plans. *Diversity and Distributions*, **14**, 462-471.

Rhodes, J. R., Tyre, A. J., Jonzen, N., McAlpine, C. A. & Possingham, H. P. (2006a). Optimizing presence-absence surveys for detecting population trends. *Journal of Wildlife Management*, **70**, 8-18.

Rhodes, J. R., Wiegand, T., McAlpine, C. A., Callaghan, J., Lunney, D., Bowen, M. & Possingham, H. P. (2006b). Modeling species' distributions to improve conservation in semiurban landscapes: Koala case study. *Conservation Biology*, **20**, 449-459.

Robertson, M. & Barker, N. (2006). A technique for evaluating species richness maps generated from collections data. *South African Journal of Science*, **102**, 78–84.

Robertson, M. P., Cumming, G. S. & Erasmus, B. F. N. (2010). Getting the most out of atlas data. *Diversity and Distributions*, **16**, 363-375.

Romo, H., García-Barros, E. & Lobo, J. M. (2006). Identifying recorder-induced geographic bias in an Iberian butterfly database. *Ecography*, **29**, 873-885.

Rondinini, C., Wilson, K. A., Boitani, L., Grantham, H. & Possingham, H. P. (2006). Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, **9**, 1136-1145.

Salzer, D. & Salafsky, N. (2006). Allocating resources between taking action, assessing status, and measuring effectiveness of conservation actions. *Natural Areas Journal*, **26**, 310-316.

Saunders, D. A., Hobbs, R. J. & Arnold, G. W. (1993). The Kellerberrin project on fragmented landscapes: A review of current information. *Biological Conservation*, **64**, 185-192.

Saunders, D. A., Hobbs, R. J. & Margules, C. R. (1991). Biological consequences of ecosystem fragmentation: A review. *Conservation Biology*, **5**, 18-32.

Schmeller, D. S., Henry, P. Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Deri, E., Budrys, E., Kull, T., Tali, K., Bauch, B., Settele, J., Van Swaay, C., Kobler, A., Babij, V., Papastergiadou, E. & Henle, K. (2009). Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conservation Biology*, **23**, 307-316.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, **24**, 467-471.

Sparks, T. H., Huber, K. & Tryjanowski, P. (2008). Something for the weekend? Examining the bias in avian phenological recording. *International Journal of Biometeorology*, **52**, 505-510.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, **142**, 2282-2292.

Szabo, J. K., Davy, P. J., Hooper, M. J. & Astheimer, L. B. (2007). Predicting spatio-temporal distribution for eastern Australian birds using Birds Australia's Atlas data: survey method, habitat and seasonal effects. *Emu* **107**, 89-99.

Szabo, J. K., Vesk, P. A., Baxter, P. W. J. & Possingham, H. P. (2011). Paying the extinction debt: woodland birds in the Mount Lofty Ranges, South Australia. *Emu*, **111**, 59–70.

Telfer, M. G., Preston, C. D. & Rothery, P. (2002). A general method for measuring relative change in range size from biological atlas data. *Biological Conservation*, **107**, 99-109.

Tulloch, A., Possingham, H. P. & Wilson, K. (2011). Wise selection of an indicator for monitoring the success of management actions. *Biological Conservation*, **144**, 141-154.

Tulloch, A. I. T. & Szabo, J. K. (accepted). Understanding the ecology and behaviour of volunteer bird surveyors. *Emu*.

Underhill, L. G., Oatley, T. B. & Harrison, J. A. (1991). The role of large scale data collection projects in the study of southern African birds. *Ostrich* **62**, 124–148.

Walsh, J. C., Wilson, K. A., Benshemesh, J., Possingham, H. P., Katzner, T. & Rondeau, D. (2012). Unexpected outcomes of invasive predator control: the importance of evaluating conservation management actions. *Animal Conservation*, n/a-n/a.

Weston, M., Fendley, M., Jewell, R., Satchell, M. & Tzaros, C. (2003). Volunteers in bird conservation: Insights from the Australian Threatened Bird Network. *Ecological Management & Restoration*, **4**, 205-211.

Weston, M. A., Silcocks, A., Tzaros, C. L. & Ingwersen, D. (2006). A survey of contributors Australian bird atlassing project: demography, skills and motivation. *Australian Journal on Volunteering*, **11**, 51-58.

Williams, P. H., Prance, G. T., Humphries, C. J. & Edwards, K. S. (1996). Promise and problems in applying quantitative complementary areas for representing the diversity of some Neotropical plants (families Dichapetalaceae, Lecythidaceae, Garyocaraceae, Ghrysobalanaceae and Proteaceae). *Biological Journal of the Linnean Society*, **58**, 125-157.

Yoccoz, N. G., Nichols, J. D. & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, **16**, 446-453.

Table 1. Nine broad hypotheses of volunteer motivations for surveying in particular locations, based on previous studies of bias in citizen science collected datasets. These hypotheses resulted in 20 models.

Hypothesis	Description	Global models	Explanatory variables
1) Accessibility	Volunteers survey in areas that are easy to access (Parnell *et al.* 2003; Reddy & Dávalos 2003; Romo *et al.* 2006; Szabo *et al.* 2007; Botts *et al.* 2010)	Access roads	roads + intensive land use
		Access town	towns + intensive land use
		Access Perth	Perth + intensive land use
		Urban	towns
		Roads	roads
2) Aesthetics	Volunteers survey in interesting landscapes (Romo *et al.* 2006)	Aesthetics	habitat diversity + protected areas+ coast + creeks
3) Expectations	Volunteers survey where they expect the birds to be in nature (near shelter, trees and water)	Expectations	distance water + distance coast + remnant veg + water + forestry
		Trees	remnant veg + forestry
		Water	distance water + distance coast + water
4) Information	Volunteers survey in areas where there is information provided to them, e.g. near visitor centres, or protected areas with information on the internet (Küper *et al.* 2006; Boakes *et al.* 2010)	Info	Perth + visitor centres + DEC web + DEC named
5) Enjoyment/ Recreation ('Fun')	Volunteers survey to have fun and enjoy the outdoors, e.g. picnic areas, waterholes, trails, parks (Küper *et al.* 2006; Boakes *et al.* 2010)	Fun	protected areas + recreation + water land use + DEC trails
6) Tourism	Volunteers survey in popular tourist destinations (Boakes *et al.* 2010)	Tourism roads	water + DEC trails + roads + visitor centres
		Tourism towns	water + DEC trails + towns + visitor centres
		Tourism Perth	water + DEC trails + Perth + visitor centres
7) Work	Bird surveys are for Environmental Impact Assessments for development/industry	Work	forestry + intensive land use
		Work access	forestry + intensive land use + towns
8) Threatened Species and Protected areas	Volunteer surveyors are concerned about threatened species and want to monitor them (Weston *et al.* 2003; Küper *et al.* 2006)	Conservation concern	protected areas + habitat diversity + threatened sp
		Protected areas	protected areas
		Threatened sp	threatened sp
9) Incentives	Volunteer surveyors are motivated by past incentives and organised	Incentives	urban + agriculture

	mass survey times, e.g. Birds in Backyards, Birds on Farms surveys organised by Birding Australia (Barrett 2000; Parsons & Major 2004)	

Table 2. Identity of environmental variables for each grid cell used in the modelling process (for further information on the derivation of these variables see Table S1).

Variable	Description	Type
roads	Density of roads (km)	continuous
towns	Average distance from a town (km)	continuous
Perth	Average distance from the capital city Perth (km)	continuous
intensive land use	Intensive land use present- residential, industrial, recreational, intensive animal production and horticulture (Department of Agriculture and Food 2006)*	categorical (0,1)
habitat diversity	Number of different habitat types (Beard 1980a; b)**	continuous
protected areas	Number of protected areas managed for conservation (by WA Department of Environment and Conservation)	continuous
coast	Coastline present	categorical (0,1)
creeks	Density of all watercourses (creeks and rivers) (km)	continuous
distance water	Average distance from a water body (creek, ocean, lake)	continuous
distance coast	Average distance from the coast (km)	continuous
remnant veg	Area of remnant vegetation (km^2)	continuous
water	Water land use present (reservoir/dam, lake, estuary, or river)	categorical (0,1)
forestry	Forestry land use present (plantation or production)	categorical (0,1)
visitor centres	Average distance from a tourist visitor centre (km)	continuous
DEC web	At least one protected area advertised on the internet: http://www.dec.wa.gov.au (WA Department of Environment and Conservation)	categorical (0,1)
DEC named	At least one named protected area (WA Department of Environment and Conservation)	categorical (0,1)
DEC trails	At least one trail in a protected area (WA Department of Environment and Conservation)	categorical (0,1)
urban	Urban land use present (Department of Agriculture and Food 2006)*	categorical (0,1)
recreation	Recreational land use present (e.g. local parks, gardens, cultural services) (Department of Agriculture and Food 2006)*	categorical (0,1)
agriculture	Area of dryland agricultural land use (grazing modified pastures, cropping, seasonal horticulture) (Department of Agriculture and Food 2006)*	continuous
threatened sp	At least one threatened species recorded in atlas 1998 to 2002	categorical (0,1)

*Land use mapped 1: 25 000 in urban areas, 1:100 000 in agricultural areas and 1:250 000 in pastoral zones

** Remnant habitat type mapped at 1:250,000, describing pre-cleared Western Australian vegetation types (Beard 1980a; b)

Table 3. Multi-model inference table for the multivariate analysis of survey probability in grid cells, showing number of model parameters *K*, deviance explained, corrected AIC (AICc), AIC differences (ΔAIC) and AIC weight *w*

Model	Rank	K	Deviance explained (%)	AICc	ΔAIC	*w*
Conservation concern	1	4	13.09	3735.17	0	1
Tourism roads	2	5	11.05	3824.93	89.76	0
Access roads	3	3	10.39	3849.1	113.93	0
Roads	4	2	9.87	3869.26	134.09	0
Threatened sp	5	2	9.09	3902.91	167.74	0
Aesthetics	6	5	8.87	3918.33	183.16	0
Fun	7	5	7.91	3959.7	224.53	0
Tourism Perth	8	5	7.05	3996.44	261.27	0
Info	9	5	6.43	4022.87	287.7	0
Protected areas	10	2	6.08	4032.11	296.94	0
Tourism towns	11	4	6.12	4034.51	299.34	0
Access towns	12	3	5.90	4041.76	306.59	0
Work access	13	4	5.94	4042.09	306.92	0
Access Perth	14	3	5.75	4048.2	313.03	0
Expectations	15	6	4.82	4094.06	358.89	0
Water	16	4	4.65	4097.32	362.15	0
Work	17	3	4.36	4107.7	372.53	0
Incentives	18	3	4.18	4115.69	380.52	0
Urban	19	2	4.13	4115.79	380.62	0
Null	20	1	0	4290.78	555.60	0

Table 4. Model parameters for the optimal model, conservation concern (see Table S3 for parameters of alternative models)

Covariates	Estimate	Std. Error	*z* value	Pr(>|*z*|)
intercept	-1.77	0.06	-29.84	<0.001
protected areas (standardised)	0.83	0.08	10.58	<0.001
habitat diversity (standardised)	0.54	0.08	6.72	<0.001
threatened sp	1.24	0.08	14.86	<0.001

Figure Legends

Figure 1. Distribution of surveys in the study area) in (a) the main atlassing period from 1998–2002, (b) the post-atlassing period from 2003–2007 (validation dataset), and (c) the test dataset from 2008–2011, also showing the location of the capital city, Perth.

Figure 2.Logistic regressions of survey probability and important explanatory variables (a) frequency of protected areas per grid cell (mean = 2.65 ± 0.14 SE), (b) number of different habitats per grid cell (mean 3.90 ± 0.03 SE), and (c) detection of a threatened species during the main atlas period 1998–2002, where the bar graph elements represent the frequency distributions of the variables relative to the presence or absence of a survey in a given grid cell.

Figure 3. Comparing actual P_i (applying the best-supported model to data from 2008–2011) with predicted P_i (using data from 2003–2007) to compare model predictions in different time periods and therefore assess the models predictive power for (a) sites not visited yet ($y = 0.2289x + 0.0129$, $r^2 = 0.995$, $P < 0.001$) and (b) sites visited from 2008–2011 ($y = 0.498x - 0.0235$, $r^2 = 0.97$, $P < 0.001$).

Figure 4. Expected benefits from the best-supported model describing threatened species and protected areas motivations of volunteer bird surveyors in 2003–2007, showing different objectives of conducting bird surveys for (a) equal representation priorities, (b) habitat stratification priorities, and (c) protected area priorities.

Figure 5. Expected benefits (probability of no survey * survey target) of surveying 3666 grid cells with different monitoring objectives of (a) equal representation (2 surveys per grid cell: Benefit = $2(1 – P_i)$, $r^2 = 1$), (b) habitat stratification (1 survey per habitat type in each grid cell: Benefit = $1.37(1 – P_i) + 1.77$, $r^2 = 0.04$), and (c) protected area prioritisation (1 survey per protected area in each grid cell: Benefit = $7.59(1 – P_i) – 2.73$, $r^2 = 0.02$).

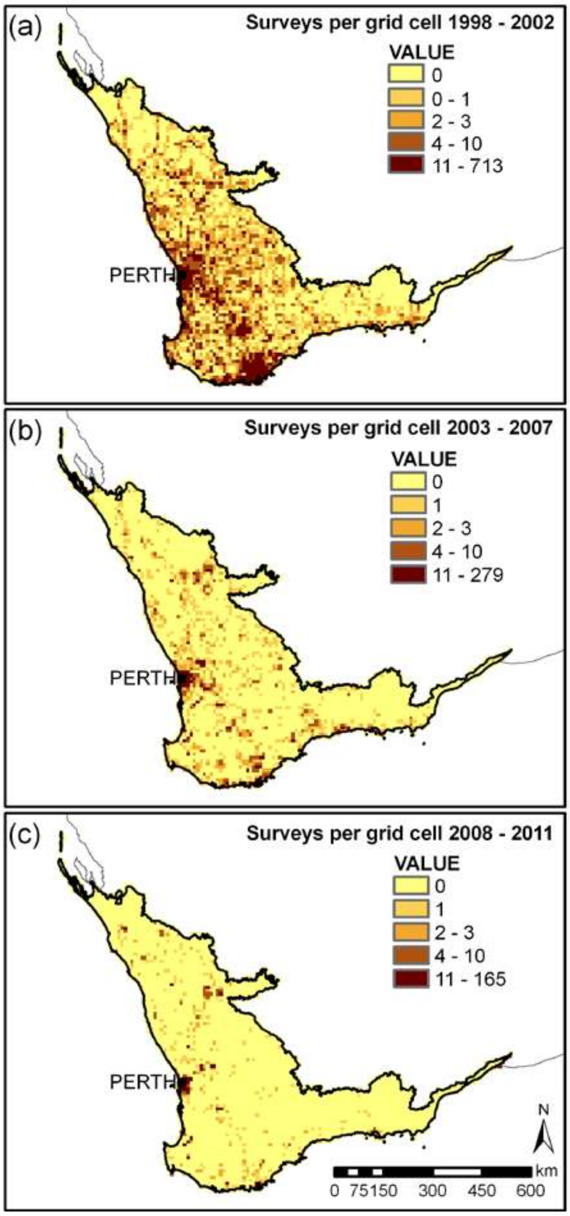

(a)
Surveys per grid cell 1998 - 2002
VALUE
0
0 - 1
2 - 3
4 - 10
11 - 713
PERTH
(b)
Surveys per grid cell 2003 - 2007
VALUE
0
1
2 - 3
4 - 10
11 - 279
PERTH
(c)
Surveys per grid cell 2008 - 2011
VALUE
0
1
2 - 3
4 - 10
11 - 165
PERTH
N
km
0 75150 300 450 600

Figure 1.

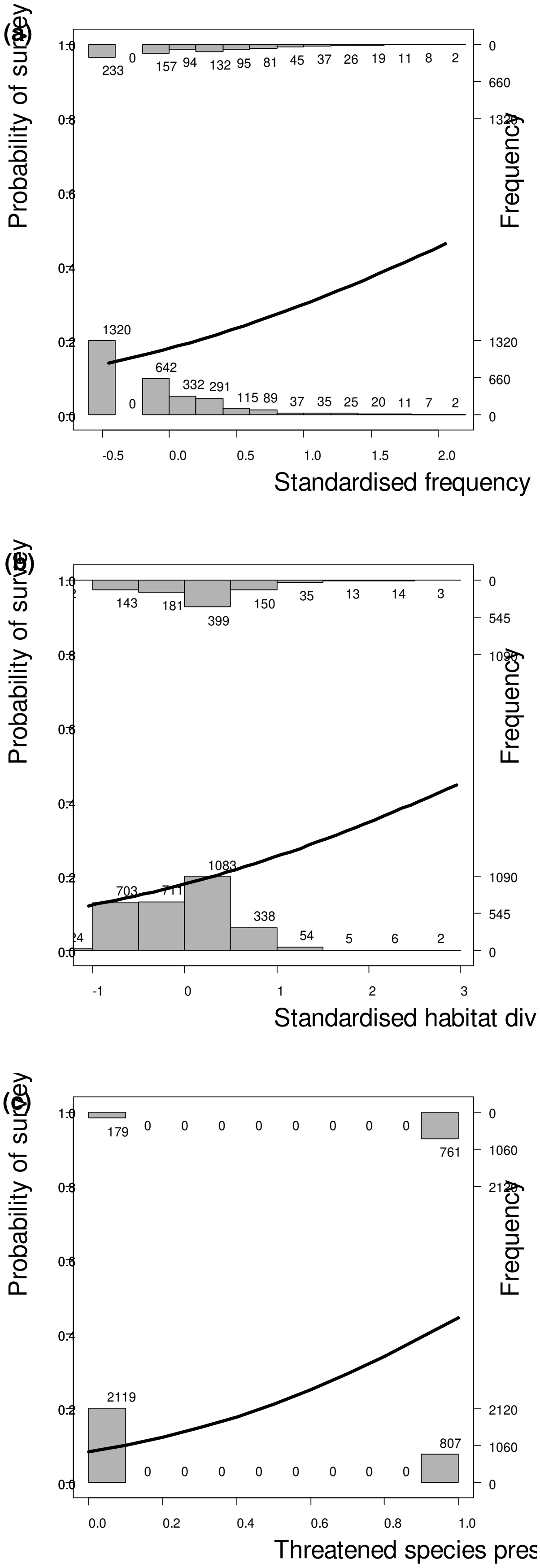
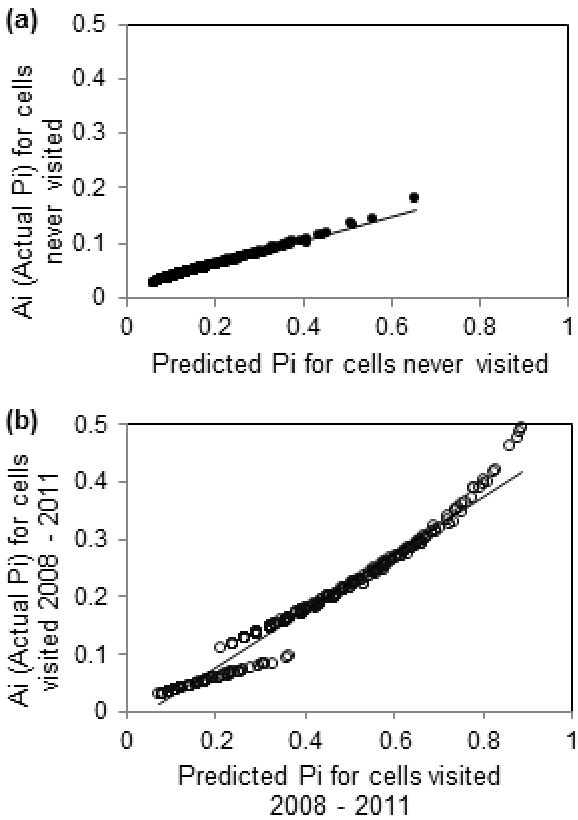

(a)
Probability of survey
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
0
660
1320
1320
660
0
233 0 157 94 132 95 81 45 37 26 19 11 8 2
1320
0 642 332 291 115 89 37 35 25 20 11 7 2
-0.5 0.0 0.5 1.0 1.5 2.0
Standardised frequency
(b)
Probability of survey
Frequency
0
545
1090
1090
545
0
143 181 399 150 35 13 14 3
24 703 711 1083 338 54 5 6 2
-1 0 1 2 3
Standardised habitat div
(c)
Probability of survey
Frequency
0
1060
2120
2120
1060
0
179 0 0 0 0 0 0 0 0 761
2119 0 0 0 0 0 0 0 0 807
0.0 0.2 0.4 0.6 0.8 1.0
Threatened species pres

Figure 2.



(a)
Ai (Actual Pi) for cells never visited
0 0.1 0.2 0.3 0.4 0.5
0 0.2 0.4 0.6 0.8 1
Predicted Pi for cells never visited
(b)
Ai (Actual Pi) for cells visited 2008 - 2011
0 0.1 0.2 0.3 0.4 0.5
0 0.2 0.4 0.6 0.8 1
Predicted Pi for cells visited 2008 - 2011

Figure 3.

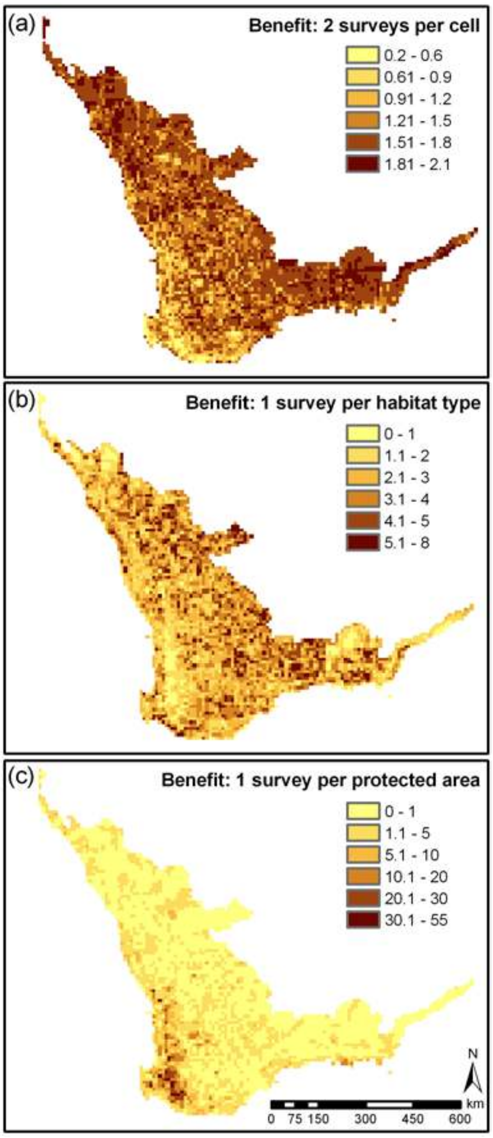
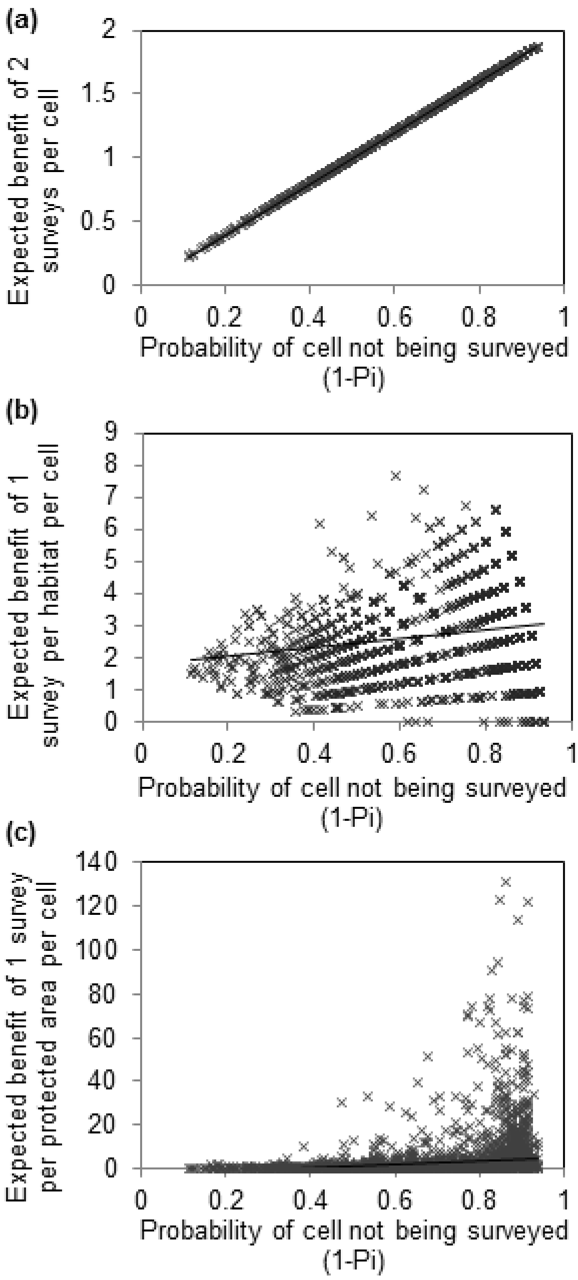

(a)
Benefit: 2 surveys per cell
0.2 - 0.6
0.61 - 0.9
0.91 - 1.2
1.21 - 1.5
1.51 - 1.8
1.81 - 2.1
(b)
Benefit: 1 survey per habitat type
0 - 1
1.1 - 2
2.1 - 3
3.1 - 4
4.1 - 5
5.1 - 8
(c)
Benefit: 1 survey per protected area
0 - 1
1.1 - 5
5.1 - 10
10.1 - 20
20.1 - 30
30.1 - 55
N
km
0 75 150 300 450 600

Figure 4.



(a)
Expected benefit of 2 surveys per cell
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
Probability of cell not being surveyed (1-Pi)
(b)
Expected benefit of 1 survey per habitat per cell
9
8
7
6
5
4
3
2
1
0
0
0.2
0.4
0.6
0.8
1
Probability of cell not being surveyed (1-Pi)
(c)
Expected benefit of 1 survey per protected area per cell
140
120
100
80
60
40
20
0
0
0.2
0.4
0.6
0.8
1
Probability of cell not being surveyed (1-Pi)

Figure 5.