Author/s:
Southwell, DM;Einoder, LD;Lahoz-Monfort, JJ;Fisher, A;Gillespie, GR;Wintle, BA

Title:
Spatially explicit power analysis for detecting occupancy trends for multiple species

Date:
2019-09-01

Citation:
Southwell, D. M., Einoder, L. D., Lahoz-Monfort, J. J., Fisher, A., Gillespie, G. R. & Wintle, B. A. (2019). Spatially explicit power analysis for detecting occupancy trends for multiple species. Ecological Applications, 29 (6), https://doi.org/10.1002/eap.1950.

Persistent Link:
https://hdl.handle.net/11343/286086

Running header: Spatially explicit power analysis

Spatially explicit power analysis for detecting occupancy trends for multiple species

Darren M. Southwell[1,*], Luke D. Einoder[2], Jose J. Lahoz-Monfort[1], Alaric Fisher[2], Graeme R. Gillespie[2], Brendan A. Wintle[1]

[1]Quantitive and Applied Ecology Group, School of BioSciences, University of Melbourne, Vic 3010, Australia

[2]Flora and Fauna Division, Department of Environment and Natural Resources, Berrimah, Darwin, Australia

[*]Corresponding Author: darren.southwell@unimelb.edu.au

Abstract

Assessing the statistical power to detect changes in wildlife populations is a crucial yet often overlooked step when designing and evaluating monitoring programs. Here, we developed a simulation framework to perform spatially explicit statistical power analysis of biological monitoring programs for detecting temporal trends in occupancy for multiple species. Using raster layers representing the spatial variation in current occupancy and species-level detectability for one or multiple observation methods, our framework simulates changes in occupancy over space and time, with the capacity to explicitly model stochastic disturbances at monitoring sites (i.e., dynamic landscapes). Once users specify the number and location of sites, the frequency and duration of surveys, and the type of detection method(s) for each species, our framework estimates power to detect occupancy trends, both across the landscape and/or within nested management units. As a case study, we evaluated power of a long-term monitoring program to detect trends in occupancy for 136 species (83 birds, 33 reptiles and 20 mammals) across and within Kakadu, Litchfield and Nitmiluk National Parks in northern Australia. We assumed continuation of an original monitoring design implemented since 1996, with the addition of camera-trapping. As expected, power to detect trends was sensitive to the direction and magnitude of the change in occupancy, detectability, initial occupancy levels, and the rarity of species. Our simulations suggest that monitoring has at least an 80% chance at detecting a 50% decline in occupancy for 22% of species across the three parks over the next 15 years. Monitoring is more likely to detect increasing occupancy trends, with at least an 80% chance at detecting a 50% increase in 87% of species. The addition of camera-trapping increased average power to detect a 50% decline in mammals compared with using only live trapping by 63%. We provide a flexible tool that can help decision-makers design and evaluate monitoring programs for hundreds of species at a time in a range of ecological settings, while explicitly considering the distribution of species and alternative sampling methods.

Keywords statistical power, spatially explicit, Kakadu, population declines, species distribution modelling, simulation, optimal monitoring, occupancy

Introduction

Monitoring the status and trends of plant and animal populations is crucial for: determining whether populations are changing over time (Gerber et al. 1999); assessing whether management strategies are working (Holling 1978); tailoring management to the current state of a population (Sewell et al. 2010), and; raising awareness and/or political support (Possingham et al. 2012). Reliably demonstrating changes in populations from biological monitoring data is notoriously difficult because monitoring programs often face financial and logistical constraints (Nichols and Williams 2006, Galvez et al. 2018), are prone to biases such as imperfect detection (Lahoz-Monfort et al. 2014), and are rarely designed with clear objectives in mind (Legg and Nagy 2006). These limitations often mean that monitoring data have little statistical power to detect population changes if they occur, leading to potential waste of resources that could be better spent elsewhere.

Power analysis is useful for designing and evaluating the likely performance of biological monitoring programs (Thomas and Juanes 1996). Statistical power is the probability that a null hypothesis of no change of interest is rejected if such a change has truly occurred (Steidl et al. 1997, Strayer 1999). It is calculated by specifying the change of interest that one wishes to detect (known as the effect size), the acceptable type 1 error rate (false alarm rate), and the 'natural' or background variation in the observed data, which is comprised of stochastic environmental variation and observation error (counting error or detection error). Power analysis can inform: 1) how likely it is that monitoring will detect important changes in a species distribution and/or abundance (Thorn et al. 2011, Loos et al. 2015); 2) the level of sampling effort required to detect an effect size with a desired degree of confidence (Barata et al. 2017), and; 3) which sampling regime will likely have the highest chance at detecting a specified level of change (Sewell et al. 2012).

In the context of biological monitoring, power analysis has primarily focused on assessing the probability of detecting changes in abundance (Rhodes et al. 2006) or occupancy (Strayer 1999, Steenweg et al. 2016, Latif et al. 2018), while accounting for observation errors such as imperfect detection (Guillera-Arroita and Lahoz-Monfort 2012). In some cases, the cost of visiting and sampling sites has been integrated with power analysis to explore trade-offs between the number of sites, and the frequency and duration of monitoring given fixed budgets and objectives (Field et al. 2005). More recently, spatially explicit power analysis has allowed for heterogeneity in occupancy/abundance to be modelled across space (Ellis et al. 2014) and the effect of site location on power to be assessed (Rhodes et al. 2006). However, despite these developments, and a range of simulation and analytical tools being freely available to aid such analyses (Bailey et al. 2007, Guillera-Arroita et al. 2010, Guillera-Arroita and Lahoz-Monfort 2012, Ellis et al. 2015), power analysis is still seldom used during the design or evaluation phase of monitoring.

Spatially explicit power analysis is particularly useful because it allows for the location, number and arrangement of monitoring sites to be explored, while accounting for spatial patterns in species distributions (Ellis et al. 2014). This is especially useful in large-scale, multi-species monitoring programs because not all species have similar distributions or tend to be found at all survey locations, resulting in significant variation in power between species for a given sampling design. The few examples of spatially explicit power analysis in the literature focus on detecting change in a single species with a single detection method, and assume that detectability of the target species remains static in time (Ellis et al. 2015, Steenweg et al. 2016). In practice, monitoring programs often seek to detect changes in occupancy and/or abundance of multiple species using a suite of detection methods across large spatial scales. At these scales, monitoring sites are often dynamic and subject to landscape-level disturbances (e.g., fire or land clearing) that might influence the distribution and detectability of species across space and time (Nimmo et al. 2014). To our knowledge, there are no simulation tools that assess power to detect trends in multiple species while explicitly incorporating the influence of spatial landscape and population processes.

In this study, we developed a simulation framework for assessing the statistical power of monitoring to detect trends in occupancy for multiple species sampled with multiple survey methods. Our goal was to provide a flexible framework that can assess the performance of a wide variety of alternative monitoring designs both across and within nested management units (e.g., state forests or national parks), with the capacity to explicitly model the influence of disturbances on occupancy and detectability. Our framework can assess detailed and realistic monitoring design options by varying the length of a monitoring program, the number and location of monitoring sites, the frequency and duration of surveys, the type(s) of detection methods, as well as the frequency of disturbances. Because it is spatially explicit and accommodates multiple species, it can also explore the effect that spatial patterns in species distributions (i.e., species rarity), initial occupancy levels, and detection rates have on power to detect occupancy trends.

Here, we demonstrate our framework by evaluating the statistical power of a large-scale vertebrate monitoring program operating across Kakadu, Litchfield and Nitmiluk National Parks in northern Australia. Initiated in 1996, this long-running program has been instrumental at documenting extensive declines in many native mammal species (Woinarski et al. 2001, Woinarski et al. 2010). However, its power to detect less dramatic trends in a broad range of mammals, reptiles and birds is thought to be low, prompting a need to evaluate its performance and revise its design. In this study, we describe how our simulation framework works and then use it to estimate power to detect trends in occupancy for 136 species over the next 15 years assuming continuation of the original monitoring design. We then demonstrate the utility of our framework by: 1) exploring the influence that rarity class has on power; 2) testing the effect that alternative site locations has on power (assuming constant monitoring effort); 3) testing the sensitivity of power to the incidence of stochastic disturbances at monitoring sites, and; 4) testing the effect of additional sampling methods (i.e., camera-trapping) on power.

Method

Occupancy and detectability raster maps

Our simulation framework for conducting spatial power analysis for detecting occupancy trends is written in the software R (R Development Core Team 2014) and is freely available on Github (https://github.com/dsouthwell/SPOTR; DOI: 10.5281/zenodo.3228292). To begin, the simulator requires occupancy and detectability raster maps for each species *s* of interest for the beginning of the simulation period ($t = 0$). Occupancy and detectability values should range from 0 to 1. Occupancy represents the probability of species *s* being present in a cell (or pixel) *j*, while detectability represents the probability of observing that species in a cell with a unit of effort for a given survey method, assuming it is present in that cell. Unique detectability maps are required for each detection method *m* for each species. For example, if a species is detected with two methods, one map of occupancy is required as well as two separate detectability maps for each sampling method. Our framework allows for any combination of up to four separate detection methods for anywhere from one to a few hundred species. All raster maps are loaded into R and manipulated using the raster package (Hijmans and van Etten 2012).

Simulations

Simulating landscape disturbances

Our simulations begin with the option of modelling stochastic disturbances in the landscape (i.e., fire or land clearing). To model stochastic disturbances, input raster maps of the disturbance history must be provided over a suitable time period (e.g., the preceding 15 years) indicating whether cells were disturbed or not. Using these layers, a 'time since disturbance' map and 'disturbance frequency' map is calculated for the landscape at the start of the monitoring period. Stochastic disturbance events are then simulated in cells for each time step of the monitoring program using a point-process model to determine if cells are either 'disturbed' or 'undisturbed'. The probability that a cell is disturbed is determined by conducting a Bernoulli trial, with the probability of success determined by a function relating the probability of a disturbance with the time since a disturbance (specified by the user). This function can be changed to influence the frequency at which disturbances occur. The 'time since last disturbance' raster map calculated at t=0 is then updated for each time step of monitoring; disturbed cells are given a value of 1, undisturbed cells are incremented by 1. The 'disturbance frequency' raster map is also updated; disturbed cells are incremented by 1, undisturbed cells retain their initial value.

Updating occupancy and detectability raster maps

If a disturbance is modelled, the occupancy and detectability raster maps are updated for each time step t using the simulated 'time since disturbance' and 'disturbance frequency' raster maps generated by the simulator. This requires specifying a statistical model for each species s that relates occupancy and detectability with the disturbance covariates (time since a disturbance and disturbance frequency), and other mapped topographic variables that might influence occupancy and detectability (e.g., terrain roughness, temperature etc) (see Einoder et al. (2018) for statistical models). Species that are sensitive to disturbances then respond either positively or negatively (in terms of occupancy and detectability) to the simulated disturbance event at a cell during each point in time.

Simulating a trend in occupancy

Our framework simulates a trend in occupancy for each species s from the start to end of a monitoring program (Figure 1). The trend can either decrease or increase over time. The magnitude of change between initial occupancy of a cell j at t=0 and occupancy at the end of the time horizon (T_{max}) is defined by the effect size E, specified by the user. For example, if the occupancy value of cell j for species s is 0.8 at the start of a monitoring program, and a 25% decline is modelled, occupancy of that cell reduces to 0.6 at T_{max}. If modelling a decline, occupancy of cell j at time t is given by:

$$\Psi_{t,s} = \Psi_{0,s}\left(1 - \left(\frac{E}{T_{max}}\right)t\right) \tag{1}$$

where E is the effect size, T_{max} is the length of the monitoring program, s is the species, and t is time. Here, the effect size is proportional to the initial occupancy value of a cell and is not the absolute change in occupancy. This means the magnitude of change depends on both the effect size and initial occupancy levels. If modelling an increasing occupancy trend, occupancy of cell j at time t is given by:

$$\Psi_{t,s} = \Psi_{0,s} + \left((1 - \Psi_{0,s})\left(\frac{E}{T_{max}}\right)t\right) \tag{2}$$

In this case, the effect size is interpreted as the proportion of the potential increase in occupancy rather than with respect to initial occupancy. For example, if initial occupancy of a cell is 0.4 and we assume a 50% increase, occupancy at T_{max} would be equal to 0.7 and not 0.6. In both cases, the effect size can be thought of as a constant 'blanket threat' on all cells because the change in occupancy is assumed to be in constant increments during each time step, acting on all cells in the same proportion. Because we rarely know *a priori* how much a population will likely change over time, a range of plausible effects sizes can be defined by the user.

Simulating the occupancy state

To determine the occupancy state of cell j at time t (i.e., whether species s is present or absent), the simulator generates a Bernoulli trial with the probability of success equal to the occupancy probability of that cell at time t. An occupancy state of 1 is assigned to cells where species s are deemed present, otherwise cells are assigned a value of 0 (absent). There is thus no Markovian dependence in the status of each cell over consecutive time periods; that is, we assumed species can move freely between cells each time step, and that beyond the role of enviornmental suitability, the existence of a species in a cell in time t does not influence the probability it will be in the same cell at time t+1.

Correcting the effect size for disturbances

The occupancy state of a cell j for species s at time t depends on the initial occupancy probability value of the cell and the effect size. However, if disturbances are modelled there will be an additional effect on occupancy for disturbance-sensitive species. Species that respond negatively to recent disturbances suffer a further reduction in occupancy, while species that respond positively will increase. The occupancy value of cell j at time t is therefore influenced by two drivers – the effect size E, which is constant across space, and the effect of a disturbance, which is a stochastic process that may act in some regions and not others. To account for these two drivers, the framework records the change in occupancy due to disturbances alone and calculates the combined effect of both processes (herafter referred to as the 'combined effect size'). For example, if a species declines by 50% but disturbances reduce occupancy by a further 10% on average across cells, the combined effect size is recorded as 60%.

Monitoring design

Once a trend in occupancy is modelled, the framework requires information on the number of monitoring sites x, the location of these sites, the time steps when monitoring occurs f (i.e., the monitoring frequency), the number of repeat visits to a site in a survey period k, and the detection method(s) *m* used to survey each species. Because occupancy and detectability is spatially explicit, sites can be arranged across the landscape in a variety of ways, including:

1) at fixed locations (e.g., representing existing or known monitoring sites);
2) randomly selected at the start of each simulation;
3) randomly stratified across environmental layers at the start of each simulation, or;
4) fixed on cells with the highest relative species richness, calculated by summing the occupancy value of each cell across species.

Detection histories

Our framework simulates detection histories at monitoring sites for each species *s* and survey method(s) *m*, assuming *k* repeat visits to a site within a surveyed time step *f* (Figure 1). The number of repeat visits to all sites must be greater than 1. To construct detection histories, a Bernoulli trial is conducted with the probability of success equal to the detection probability for that survey method at cell *j,* given the species is present. Successful detections are assigned a 1, unsuccessful attempts a 0 for each repeat visit *k* to occupied sites. If a species is detected with more than one method, separate detection histories are generated given the unique detection probability of each method.

Estimating a trend in occupancy

Our simulator calls the R the package 'unmarked' (Fiske and Chandler 2011) to estimate occupancy for species *s* during each year of monitoring *f* using the simulated detection histories. If species can be detected with more than one method, detection histories are grouped with the method defined as an observation covariate. The effect of a trend in occupancy over time is estimated by fitting an occupancy state model to the simulated detection histories:

$$logit(\Psi_{t,s}) = \alpha_0 + \alpha_1 \times time \tag{3}$$

where Ψ is occupancy in year *t* for species *s*, α_0 is the intercept and α_1 is the trend in occupancy. Additional covariates could be defined in the occupancy and/or detection models to investigate other management or policy needs.

Calculating power

A one or two-tailed non-zero significance test is conducted to determine if the trend parameter α_1 fitted to the simulated detection histories is statistically significant, given the Type I error rate. A two-tailed test checks whether the upper and lower confidence intervals around the trend parameter are the same sign (i.e., both positive or both negative), while a one-tailed test checks if the lower or upper confidence interval is less than or greater than zero, respectively, depending on the expected direction of the effect.

Simulations are repeated *n* times for a given effect size. Statistical power (1-β) is calculated as the proportion of times a significant trend in occupancy is detected from the simulated detection histories. Power is calculated across all sites (referred to as landscape-level power), and/or within nested management units (refererd to as park-level power), if these are defined by the user. Nested management units could include multiple parks or reserves within the broader landscape. A diagram of the simulation framework is presented in Figure 1.

Case study: Three Parks Monitoring Program

To demonstrate our simulation framework, we estimated the power of a large-scale monitoring progam to detect occupancy trends in vertebrate species across and within Kakadu, Nitmiluk and Litchfield National Parks in northern Australia (Figure 2a). This monitoring program (hereafter referred to as the 'Three Parks Program') has operated since 1996, detecting over 246 bird, mammal and reptile species at 241 sites. Sites (100 x 100m, with an internal 50 x 50m trapping grid) have been surveyed every 5-6 years using a standard method of live trapping (pit, cage and elliot trapping), active searches and spotlighting (see Woinarski et al. (2001) and Appendix S1, S3). Camera traps have also been added in recent years following the method of Gillespie et al. (2015).

The program has been instrumental at detecting declines in small to medium-bodied mammals (Woinarski et al. 2012), but is currently being evaluated to address concerns about its design (e.g., site locations and timing of surveys), and low detectability of many species (Einoder et al. 2018). As the first step in the evaluation process, we estimated power to detect occupancy trends in species assuming a continuation of the original monitoring protocol, with the addition of camera traps. To further demonstrate our simulator, we explored the effect of alternative site placements and stochastic fire disturbances on power.

Occupancy, detectability and covariate raster maps

To begin our simulations, we obtained occupancy and detectability raster maps for 136 species (20 mammals, 83 birds, 33 reptiles) recorded during the Three Parks Program (Einoder et al. 2018). We were restricted to this set of species because Einoder et al. (2018) found that there were too few detections to adequately fit occupancy-detection models to the remaining species. Developing additional models using alternative methods and/or data was beyond the scope of this study. Detectability for the 136 species was estimated during one day/night for live trapping methods, spotlighting and active searches and for one week of camera trapping. All raster maps were clipped to the three parks at 1 km resolution (Appendix S2).

To explore the effect of occupancy patterns on power, we assigned species into four rarity classes based on the predicted occupancy maps: 1) widespread distribution, high occupancy (63 species); 2) widespread distribution, low occupancy (40 species); 3) fragmented distribution, high occupancy (25 species), and; 4) fragmented distribution, low occupancy (8 species) (see Appendix S3: Table S1). These categories are similar to the rarity classes defined by Rabinowitz (1981) but do not include explicit information on local densities within habitat. We assumed species with a mean occupancy value of less than 0.1 had low occupancy; species above this threshold were assumed to have high occupancy. This threshold was selected based on similar rarity categorisations by Wheeler (1988) and Prendergast et al. (1993). We categorised species distributions as either widespread or fragmented by visually inspecting occupancy maps and by consulting experts with knowledge of the species and study region.

Monitoring design

We assessed power to detect occupancy trends for the 136 species in the next 15 years (T_{max}) when 241 sites (148 in Kakadu, 52 in Nitmiluk, 41 in Litchfield) are monitored every 5 years for 3 nights (Figure 2a). We first assumed a static landscape over time with no disturbances. To demonstrate the effect of site locations on power, we ran three additional scenarios: 1) sites were positioned randomly each simulation (Figure 2b); 2) sites were randomly stratified across three broad vegetation types (lowland woodland, lowland open forest, sandstone woodland; Figure 2c), and; 4) sites were positioned on cells with the highest relative species richness (Figure 2d).

We assumed birds were surveyed with two methods (active searches and/or spotlighting); reptiles were surveyed with two methods (pit fall traps and/or spotlighting); and mammals were surveyed with up to four methods (any combination of spotlighting, pit traps, Elliott traps, cage traps and/or camera traps) (Appendix S3: Table S1). We assumed camera traps were deployed for 5 weeks to detect mammals only, but also ran a scenario without cameras to explore their contribution to power.

Simulating stochastic disturbances at monitoring sites

We simulated stochastic fire disturbances at monitoring sites and adjusted the effect size for fire-sensitive species. To model the incidence of fire, we obtained fire history raster maps for the preceding 15 years (2004–2014) from the North Australia and Rangelands Fire Information (NAFI) website (http://www.firenorth.org.au/nafi2/) and simulated the incidence of fire at sites using the hazard function (i.e., the probablity of a site burning in any given year as a function of time since fire) reported by Gill et al. (2000), resulting in cells burning on average every 2 – 3 years. Occupancy and detectability rasters were updated using the statistical models reported by Einoder et al. (2018) relating occupancy and detectability to fire, topographic and climatic covariates (Appendix S2: Figure S1). The effect of fire on the effect size and power is presented for a selection of species with low, moderate and high power.

Simulations

We ran all scenarios for a range of effect sizes (10%, 30%, 50%, 70%, 90%) assuming both increasing and decreasing trends, with 1000 simulations (*n*) run for each combination of species and effect size. We conducted a two-tailed test with a type I error rate of α=0.05, and calculated power across and within the three parks.

Results

Power to detect occupancy trends

Our simulations suggest that continuing with the original monitoring design (241 sites) has sufficient power ($1-\beta>0.8$) to detect a 70% decline in occupancy in 46% of the species modelled (36 birds, 15 reptiles and 11 mammals) across the three parks over the next 15 years (Figure 3a,b,c) (Appendix S4: Table S1). As expected, power decreased as the effect size decreased, with at least an 80% chance at detecting a 50% decline for 22% of modelled species (18 birds, 8 reptiles and 4 mammals) across the three parks. Monitoring is unlikely to detect 10% declines in occupancy for any of the species with greater than 80% power.

Power to detect increasing trends in occupancy for the species modelled was considerably higher than decreasing trends. Increasing occupancy trends with an effect size of 70% were detected in 88% of modelled species (74 birds, 29 reptiles and 17 mammals) with at least 80% power (Figure 4), 50% increases were detected in 87% of modelled species (72 birds, 29 reptiles and 17 mammals) with sufficient power, while 10% increases were detected in 21% of modelled species (17 birds, 6 reptiles and 6 mammals) (Appendix S4: Table S2).

Landscape versus park-level power

At a park level, power was highest in Kakadu due to there being more sites in this park (Figure 3a). There was at least an 80% chance of detecting a 70% decline in occupancy for 31% of modelled species (26 birds, 12 reptiles and 5 mammals) in Kakadu, 12% of modelled species in Litchfield (11 birds, 4 reptiles, 2 mammals) and 11% of modelled species in Nitmiluk (10 birds, 4 reptiles, 1 mammals) (Figure 3d-l). In contrast, monitoring could confidently detect a 70% increasing trend in 85% of modelled species in Kakadu, 52% of modelled species in Litchfield and 64% of modelled species in Nitmiluk (Figure 4d-l).

The influence of site location on power

Power to detect occupancy trends was sensitive to the spatial arrangement of sites within the three parks (Figure 5a); however, no one placement of sites performed best for all species. Targeting monitoring towards cells with the highest expected species richness maximised power for 41% of modelled species, assuming a 50% decline in occupancy. Continuing to monitor at existing sites was best in terms of power for 30% of modelled species, and was a better approach than employing a random selection of sites (14% of modelled species) or randomly stratifying sites across the three broad vegetation types (14% of modelled species). The sensitivity of power to the alternative site placements is presented for three example species in Figure 5a.

The influence of rarity class on power

Monitoring is most likely to detect declines in widely distributed species with high initial occupancy values (rarity class 1; Figure 3) because these species are most likely available for detection at monitoring sites. Power was lowest for species with low initial occupancy values (rarity classes 2 and 4), because these species were mostly unavailable for detection, even if detectability was high. Species with fragmented distributions, but high initial occupancy values (rarity class 3) varied considerably in terms of power, depending on whether monitoring sites aligned with patches of high occupancy (Figure 3). However, we note that there was considerable variation in power within each rarity class. This is because power depends on both occupancy and detectability, so a widespread species with high occupancy could still have low power if detectability was low.

The influence of fire disturbances and camera-trapping on power

Simulating fire disturbances at sites influenced the effect size and power to detect trends for fire-sensitive species (Figure 5b). The sensitivity of power to fire depended on the hazard function (i.e., the probability of cells burning given time since fire) and the direction and magnitude of the effect of fire on occupancy and detectability. Similarly, the addition of camera-trapping increased power to detect occupancy trends in medium-to-large mammals (Figure 5c). For example, the average power to detect a 50% decline in occupancy for mammals was 0.36 with cameras and 0.22 with only live trapping, a gain of 63%. This increase in power is because the cumulative probability of detection over the sampling period of 5 weeks was relatively high compared to the cumulative probability of detection during 3 nights of live trapping.

Discussion

We developed a simulation framework for estimating power to detect occupancy trends for multiple species in dynamic landscapes. Our framework utilises occupancy and detectability raster maps for target species, allowing decision-makers to test the performance of monitoring sites positioned at any location in the landscape. It builds on existing applications of spatially explicit power analysis (Ellis et al. 2015) by: 1) simulating either increasing or decreasing occupancy trends for multiple species; 2) allowing for multiple detection methods at sites (and combining those data in a coherent way), 3) providing the option of explicitly model stochastic disturbances (e.g., fire or vegetation loss), and; 4) comparing the effect of site placement on power. These additional features allow for a more realistic and flexible set of monitoring scenarios to be considered by decision-makers when evaluating or designing large-scale, multi-species monitoring programs.

We demonstrated the utility of our framework by evaluating alternative monitoring designs for a realistically complex and large monitoring program in northern Australia. Our simulations suggest that continuing with the original Three Parks Program will likely detect 50% declines in one fifth (22%) of the species modelled. We also found that monitoring is more likely to detect increasing occupancy trends. This difference in power between increasing and decreasing trends was due to our interpretation of the effect size. When simulating a decline, we assumed the effect size was proportional to initial occupancy, whereas the effect size was proportional to the potential increase in occupancy for increasing trends. This meant the magnitude of change was often much greater when modelling an increasing trend because many species had very low initial occupancy values. Our framework could easily be modified so that the effect size reflects an absolute change in occupancy rather than proportional change; however, final occupancy would have to be truncated at either zero or one in many cases. For both increasing and decreasing trends, the proportion of species we could detect changes in was with respect to the 136 species modelled by Einoder et al. (2018) and not the 247 number species recorded at least once during the Three Parks Program. We expect power for the remaining 111 species not included here to be very low given that Einoder et al. (2018) could not adequately fit occupancy-detection models to available data because of too few detections.

Our framework generated species-level power curves both across and within nested reserves for the Three Parks Program. Such information provides decision-makers with guidance as to the likely chance at detecting changes in species distributions across different spatial scales. However, whether the Three Parks Program is effective depends on the fundamental objectives of monitoring, the desired level of power and the acceptable alpha level (or false alarm rate) (Possingham et al. 2012). For example, if the goal of monitoring is to detect changes in the most widespread and common species at a landscape scale, then the existing Three Parks Program might be considered adequate at detecting change. Alternatively, if the goal is to detect changes in only the rarest and/or most cryptic species with a high level of confidence, or to maximise the number of species in which a change can be detected, our results might prompt managers to re-consider the available budget or allocation of monitoring effort. An interesting area of further research would be to investigate whether monitoring more sites within the three reserves for a shorter amount of time (or vice-versa) results in an increase in power (Field et al. 2005, Mackenzie and Royle 2005). Such analyses could be easily conducted with our framework by comparing power estimates for target species by re-running simulations for alternative monitoring scenarios.

A novel feature of our framework is that any unique combination of up to four sampling methods can be defined for each species. This is particularly useful for evaluating or designing large-scale monitoring programs where more than one sampling method might be deployed at sites to detect species. The benefit of incorporating multiple sampling methods is that the relative contribution of each to power can be explored. For example, many mammals can be detected with both live trapping and camera trapping in the Three Parks Program (Einoder et al. 2018). We demonstrated camera trapping for 5 weeks has a significant benefit on power compared to if only live trapping over 4 days and 3 nights. This result is not surprising: cameras can remain at sites for extended periods, which can increase the chance that a species is detected at least once to confirm presence (Smith et al. 2017). The value of our framework is that it can quantify this benefit in terms of power to detect change. An interesting extension to this work would be to compare the cost-effectiveness of sampling methods; that is, weigh up the benefit of each methods in terms of power compared to the expense (Balme et al. 2009). While camera trapping might increase power for mammals, this gain should be considered in light of the extra time needed to collect, process and analyse the images.

The benefit of a spatially explicit framework is that power can be assessed for sites positioned at any location in the landscape while explicitly considering the overlapping distribution of multiple species. This allows users to test for adequate landscape-level site stratification across: 1) environmental gradients; 2) management units, such as parks or conservation reserves; and, 3) the extent and range of species distributions (Rhodes et al. 2006). Our approach therefore allows for important spatial processes to be incorporated into decision-making, which if ignored, could provide a naively optimistic view of the power of a given monitoring design option. For example, we demonstrated that monitoring should consider spatial patterns of species distributions (i.e., rarity class) and that the location and arrangement of sites influences power to detect change. Positioning sites at locations with high initial occupancy maximised our chance at detecting declines in occupancy because the absolute change in occupancy was greatest. Because the framework is spatially explicit, this means that decision-makers can make informed decisions about which sites to remove from existing programs, or where to position new sites in the landscape at previously unsampled locations.

Our framework simulates a change in species occupancy through time, but also has the capacity to simulate an additional change in cells due to stochastic disturbances. This opens the possibility of simulating disturbances in some regions of the landscape and not others as well as the frequency at which disturbances occur. Simulating stochastic disturbances requires knowledge of the disturbance history throughout the landscape and the relationship between the probability of a disturbance and the time since the last disturbance. Modelling disturbances in this way has an implicit spatial component because the probability of a disturbance depends on the disturbance history (McCarthy and Carey 2002). However, this component of our simulation tool could be further developed to model spatial processes. For example, we could modify the point process fire model used in our case study to simulate the process of ignition and spread of fire between cells using cellular automata models (Karafyllidis and Thanailakis 1997). This would potentially allow users to explore the effect of different disturbance regimes on spatial patterns of species, as well as the power of alternative monitoring designs at detecting such change. The limitations of expanding this component of the framework, however, is that it might come at considerable computational cost because the spatial processes must be repeatedly simulated every time step.

We modelled trends in occupancy rather than trends in abundance over time. Monitoring occupancy as a proxy for population size is relatively common, especially across large spatial scales, as it is generally collected with greater ease and cost-effectiveness than more detailed demographic data required to estimate population size. By simulating changes in occupancy, we assumed a 1:1 relationship between occupancy and abundance (Stanley and Royle 2005), which meant that all occupied cells changed in the same way, regardless of their occupancy probability. This ignores the number of individuals that might be present within an occupied cell. In practice, the abundance of species will vary across occupied cells, and raster cells with a high abundance will likely have less chance of becoming locally extinct than cells with only a few individuals. Local abundance at monitoring sites will also likely influence rates of detection (McCarthy et al. 2013), although the relationship between occupancy and abundance/detectability is rarely evaluated (Gaston et al. 2000). Further research into relationships between occupancy probability, effect size and detectability would benefit the literature and improve our simulation framework, although we note that such relationships are likely species-specific.

Our current implementation makes simplifying assumptions that could be relaxed to provide an appropriate level of realism or to utilise data where it exists. For example, we assumed occupancy changed at a constant rate across space and time, was static within years, and was constant across space (ignoring disturbances), regardless of the initial occupancy value of a cell. Our framework could be expanded to model different patterns of change across space (i.e., range contractions, expansions or shifts) (Steenweg et al. 2016) and time (i.e., linear versus exponential declines). We also modelled declines in individual species rather than in species richness. Simulating declines in species richness would have required additional assumptions about how species decline relatively to one another. Finally, our simulation tool requires initial raster maps of occupancy and detectability at the start of simulations. When initial presence-absence data are not available, expert opinion on habitat preferences, detection rates, and initial distributions can be developed in the form of habitat suitability indices (Burgman et al. 2001) or inferred from similar or related species. However, we emphasise that simulations become more realistic for species with accurate occupancy and detectability models and maps, preferably based on strong biological survey data.

Conclusion

We provide a framework that decision-makers can use to assess and compare alternative monitoring designs at detecting occupancy trends. Our tool is flexible enough to accommodate any combination of up to four survey methods to detect either decreasing or increasing occupancy trends for hundreds of species. We demonstrate its use by estimating power of a long-running monitoring program to detect trends in vertebrates in northern Australia. Our results are now being used by managers to re-evaluate the monitoring objectives and optimal allocation of effort. Although we present one case study, our framework could be applied across a range of ecological settings, including terrestrial and marine ecosystems. Incorporating spatially explicit power analysis into conservation planning will result in more robust monitoring, and ultimately lead to more confident and earlier detection and reporting of population changes when they occur.

Acknowledgements

This study was supported by funding from the Australian Government's National Environmental Science Programme through the Threatened Species Recovery Hub. We thank Anja Skroblin and David Wilkinson for useful comments on the manuscript. Monitoring was assisted by financial support from LTERN. Charmaine Redmond, Terry Marney, Alys Stevens, Jenni Risler, Brydie Hill, Dani Stokeld, Tony Popic, Tim Gentles, Kathryn Buckley, Emmanuelle Corolleur, Nicholas Cuff and Matthew Fegan all contributed to data collection and processing.

Literature cited

Bailey, L. L., J. E. Hines, J. D. Nichols, and D. I. MacKenzie. 2007. Sampling design trade-offs in occupancy studies with imperfect detection: Examples and software. Ecological Applications **17**:281-290.

Balme, G. A., L. T. B. Hunter, and R. Slotow. 2009. Evaluating Methods for Counting Cryptic Carnivores. Journal of Wildlife Management **73**:433-441.

Barata, I. M., R. A. Griffiths, and M. S. Ridout. 2017. The power of monitoring: optimizing survey designs to detect occupancy changes in a rare amphibian population. Scientific Reports **7**.

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

Einoder, L., D. Southwell, J. Lahoz-Monfort, G. Gillespie, and B. Wintle. 2018. Occupancy and detectability modelling of vertebrates in northern Australia using multiple sampling methods. Plos One **13(9): e0203304**:https://doi.org/10.1371/journal.pone.0203304.

Ellis, M. M., J. S. Ivan, and M. K. Schwartz. 2014. Spatially Explicit Power Analyses for Occupancy-Based Monitoring of Wolverine in the U.S. Rocky Mountains. Conservation Biology **28**:52-62.

Ellis, M. M., J. S. Ivan, J. M. Tucker, and M. K. Schwartz. 2015. rSPACE: Spatially based power analysis for conservation and ecology. Methods in Ecology and Evolution **6**:621-625.

Field, S. A., A. J. Tyre, and H. P. Possingham. 2005. Optimizing allocation of monitoring effort under economic and observational constraints. Journal of Wildlife Management **69**:473-482.

Fiske, I., and R. Chandler. 2011. unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. Journal of Statistical Software **43**:1-23.

Galvez, N., G. Guillera-Arroita, F. A. V. St John, E. Schuttler, D. W. Macdonald, and Z. G. Davies. 2018. A spatially integrated framework for assessing socioecological drivers of carnivore decline. Journal of Applied Ecology **55**:1393-1405.

Gaston, K. J., T. M. Blackburn, J. J. D. Greenwood, R. D. Gregory, R. M. Quinn, and J. H. Lawton. 2000. Abundance-occupancy relationships. Journal of Applied Ecology **37**:39-59.

Gerber, L. R., D. P. DeMaster, and P. M. Kareiva. 1999. Gray whales and the value of monitoring data in implementing the US Endangered Species Act. Conservation Biology **13**:1215-1219.

Gill, A. M., P. G. Ryan, P. H. R. Moore, and M. Gibson. 2000. Fire regimes of World Heritage Kakadu National Park, Australia. Austral Ecology **25**:616-625.

Gillespie, G. R., K. Brennan, T. Gentles, B. Hill, J. Low Choy, T. Mahney, A. Stevens, and D. Stokeld. 2015. A guide for the use of remote cameras for wildlife surveys in northern Australia. National Environmental Research Program, Northern Australia Hub, Charles Darwin University, Casuarina, NT.

Guillera-Arroita, G., and J. J. Lahoz-Monfort. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection. Methods in Ecology and Evolution **3**:860-869.

Guillera-Arroita, G., M. S. Ridout, and B. J. T. Morgan. 2010. Design of occupancy studies with imperfect detection. Methods in Ecology and Evolution **1**:131-139.

Hijmans, R. J., and J. van Etten. 2012. raster: Geographic analysis and modelling with raster data. http://CRAN.R-project.org/package=raster

Holling, C. S. 1978. Adaptive environmental assessment and management. John Wiley and Sons, London, UK.

Karafyllidis, I., and A. Thanailakis. 1997. A model for predicting forest fire spreading using cellular automata. Ecological Modelling **99**:87-97.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. A. Wintle. 2014. Imperfect detection impacts the performance of species distribution models. Global Ecology and Biogeography **23**:504-515.

Latif, Q. S., M. M. Ellis, V. A. Saab, and K. Mellen-McLean. 2018. Simulations inform design of regional occupancy-based monitoring for a sparsely distributed, territorial species. Ecology and Evolution **8**:1171-1185.

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

Loos, J., J. Hanspach, H. von Wehrden, M. Beldean, C. I. Moga, and J. Fischer. 2015. Developing robust field survey protocols in landscape ecology: a case study on birds, plants and butterflies. Biodiversity and Conservation **24**:33-46.

Mackenzie, D. I., and J. A. Royle. 2005. Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology **42**:1105-1114.

McCarthy, M. A., and G. A. Carey. 2002. Fire regimes in landscapes: models and realities. Pages 77-93 *in* R. Bradstock, J. Williams, and M. Gill, editors. Flammable Australia: the fire regimes and biodiversity of a continent. Cambridge University Press, Cambridge.

McCarthy, M. A., J. L. Moore, W. K. Morris, K. M. Parris, G. E. Garrard, P. A. Vesk, L. Rumpff, K. M. Giljohann, J. S. Camac, S. S. Bau, T. Friend, B. Harrison, and B. Yue. 2013. The influence of abundance on detectability. Oikos **122**:717-726.

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. Trends in Ecology & Evolution **21**:668-673.

Nimmo, D. G., L. T. Kelly, L. M. Farnsworth, S. J. Watson, and A. F. Bennett. 2014. Why do some species have geographically varying responses to fire history? Ecography **37**:805-813.

Possingham, H. P., B. A. Wintle, R. A. Fuller, and L. N. Joseph. 2012. The Conservation return on investment from ecological monitoring. Pages 49-61 *in* D. B. Lindenmayer and P. Gibbons, editors. Making Biodiversity Monitoring Happen in Australia. CSIRO Publishing, Melbourne.

Prendergast, J. R., R. M. Quinn, J. H. Lawton, B. C. Eversham, and D. W. Gibbons. 1993. Rare species, the coincidence of diversity hotspots and conservation strategies. Nature **365**:335-337.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabinowitz, D. 1981. Seven forms of rarity. Pages 205-217 *in* H. Synge, editor. The biological aspects of rare plant conservation. John Wiley & Sons, Chichester.

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

Sewell, D., T. J. C. Beebee, and R. A. Griffiths. 2010. Optimising biodiversity assessments by volunteers: The application of occupancy modelling to large-scale amphibian surveys. Biological Conservation **143**:2102-2110.

Sewell, D., G. Guillera-Arroita, R. A. Griffiths, and T. J. C. Beebee. 2012. When Is a Species Declining? Optimizing Survey Effort to Detect Population Changes in Reptiles. Plos One **7**.

Smith, J., S. Legge, A. James, and K. Tuft. 2017. Optimising camera trap deployment design across multiple sites for species inventory surveys. Pacific conservation biology **23**:43-51.

Stanley, T. R., and J. A. Royle. 2005. Estimating site occupancy and abundance using indirect detection indices. Journal of Wildlife Management **69**:874-883.

Steenweg, R., J. Whittington, M. Hebblewhite, A. Forshner, B. Johnston, D. Petersen, B. Shepherd, and P. M. Lukacs. 2016. Camera-based occupancy monitoring at large scales: Power to detect trends in grizzly bears across the Canadian Rockies. Biological Conservation **201**:192-200.

Steidl, R. J., J. P. Hayes, and E. Schauber. 1997. Statistical power analysis in wildlife research. Journal of Wildlife Management **61**:270-279.

Strayer, D. L. 1999. Statistical power of presence-absence data to detect population declines. Conservation Biology **13**:1034-1038.

Thomas, L., and F. Juanes. 1996. The importance of statistical power analysis: An example from Animal Behaviour. Animal Behaviour **52**:856-859.

Thorn, M., M. Green, P. W. Bateman, S. Waite, and D. M. Scott. 2011. Brown hyaenas on roads: Estimating carnivore occupancy and abundance using spatially auto-correlated sign survey replicates. Biological Conservation **144**:1799-1807.

Wheeler, B. D. 1988. Species richness, species rarity and conservation evaluation of rich-fen vegetation in lowland England and wales. Journal of Applied Ecology **25**:331-352.

Woinarski, J. C. Z., M. Armstrong, K. Brennan, A. Fisher, A. D. Griffiths, B. Hill, D. J. Milne, C. Palmer, S. Ward, M. Watson, S. Winderlich, and S. Young. 2010. Monitoring indicates rapid and severe decline of native small mammals in Kakadu National Park, northern Australia. Wildlife Research **37**:116-126.

Woinarski, J. C. Z., A. Fisher, M. Armstrong, K. Brennan, A. D. Griffiths, B. Hill, J. L. Choy, D. Milne, A. Stewart, S. Young, S. Ward, S. Winderlich, and M. Ziembicki. 2012. Monitoring indicates greater resilience for birds than for mammals in Kakadu National Park, northern Australia. Wildlife Research **39**:397-407.

Woinarski, J. C. Z., D. J. Milne, and G. Wanganeen. 2001. Changes in mammal populations in relatively intact landscapes of Kakadu National Park, Northern Territory, Australia. Austral Ecology **26**:360-370.

Data availability: The R code for running the simulation tool is freely available on Zenodo: http://doi.org/10.5281/zenodo.3228292

Figure 1: Structure of the spatially explicit power analysis framework for multiple species in dynamic landscapes.

Figure 2: Map of the study region including: a) the location of 241 existing monitoring sites in Kakadu, Litchfield and Nitmiluk National parks; b) example of sites selected randomly across the parks; c) example of sites stratified randomly across three broad vegetation types - open forest (dark grey), open woodland (light grey) and sandstone woodland (grey); d) sites positioned on cells with the highest relative species richness. The black dot on the insert map in panel b shows the location of the study site in Australia.

Figure 3: Statistical power (y-axis) to detect declines in occupancy (x-axis) for birds (left column), reptiles (middle column) and mammals (right column) over a 15 year monitoring program across the three parks (top row) and within Kakadu, Litchfield and Nitmiluk parks in northern Australia. We assumed 241 sites are monitored every 5 years, with live trapping for 3 nights and camera-trapping for 5 weeks. This scenario assumes no disturbances. Orange lines indicate species in rarity class 1 (widespread distribution, high occupancy), blue lines rarity class 2 (widespread distribution, low occupancy), grey lines rarity class 3 (fragmented distribution, high occupancy), green lines rarity class 4 (fragmented distribution, low occupancy). Horizontal dashed lines show 80% power.

Figure 4: Statistical power (y-axis) to detect increasing trends in occupancy (x-axis) for birds (left column), reptiles (middle column) and mammals (right column) over a 15 year monitoring program across (top row) and within Kakadu, Litchfield and Nitmiluk parks in northern Australia. We assumed 241 sites are monitored every 5 years, with live trapping for 3 nights and camera-trapping for 5 weeks. See Figure 3 for colour references.

Figure 5: Statistical power (y-axis) to detect declines in occupancy (x-axis) over 15 years for: a) the Arafura Fantail (*Rhipidura dryas*), Brown Falcon (*Falcon berigora*), and the Blue-faced Honeyeater (*Entomyzon cyanotis*), assuming 241 sites are positioned on cells with the highest relative species richness (black line), randomly positioned (dark blue line), randomly stratified across three broad vegetation types (light blue line), and positioned at existing sites (grey line); b) the Bynoe's gecko (*Heteronotia binoei*), Eastern striped skink (*Ctenotus robustus*), North-eastern Orange-tailed Slider (*Lerista orientalis*) when fire is simulated (black solid line) and when it is not simulated (black dashed line) at monitoring sites, and; c) the Common Rock Rat (*Zyzomys argurus*), Northern Brown Bandicoot (*Isoodon macrourus*) and Red-cheeked Dunnart (*Sminthopsis virginiae*) with (solid lines) and without (dashed lines) camera-trapping.

684

Figure 1

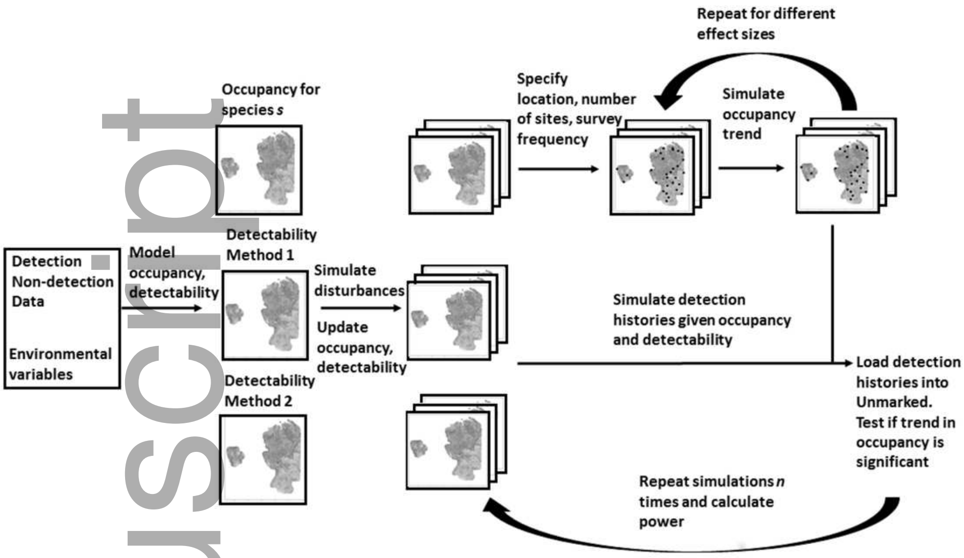

Repeat for different effect sizes
Occupancy for species s
Specify location, number of sites, survey frequency
Simulate occupancy trend
Detection
Non-detection
Data
Environmental variables
Model occupancy, detectability
Detectability Method 1
Simulate disturbances
Update occupancy, detectability
Simulate detection histories given occupancy and detectability
Load detection histories into Unmarked. Test if trend in occupancy is significant
Detectability Method 2
Repeat simulations n times and calculate power

Figure 2

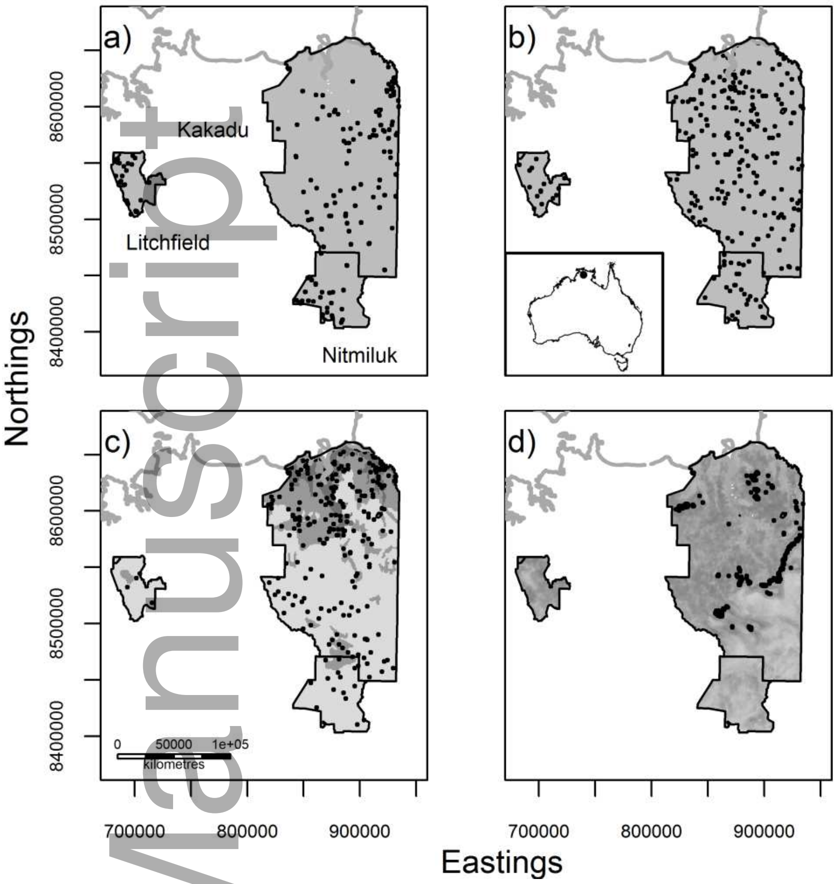
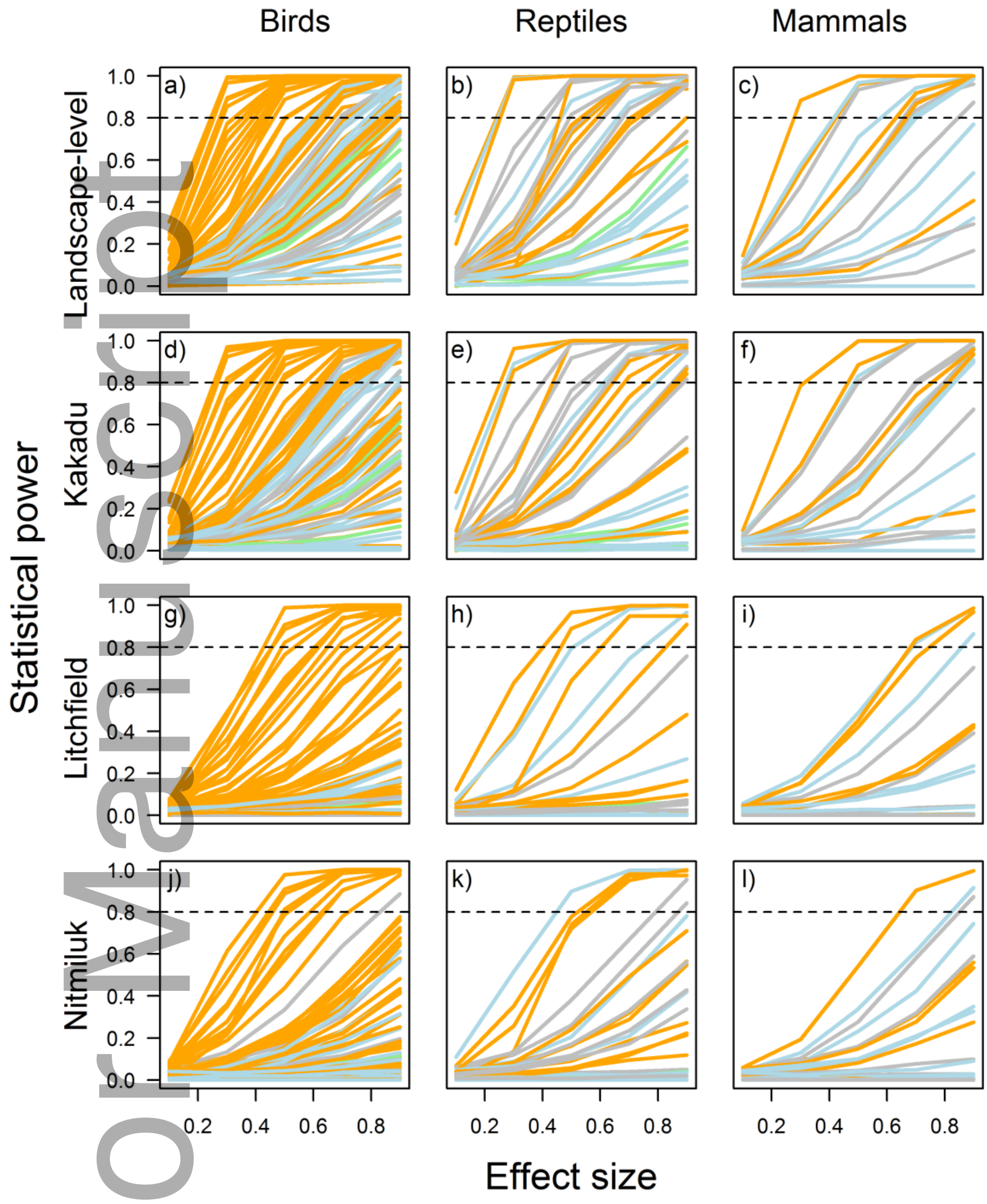


Figure 3



692

693

694 Figure 4

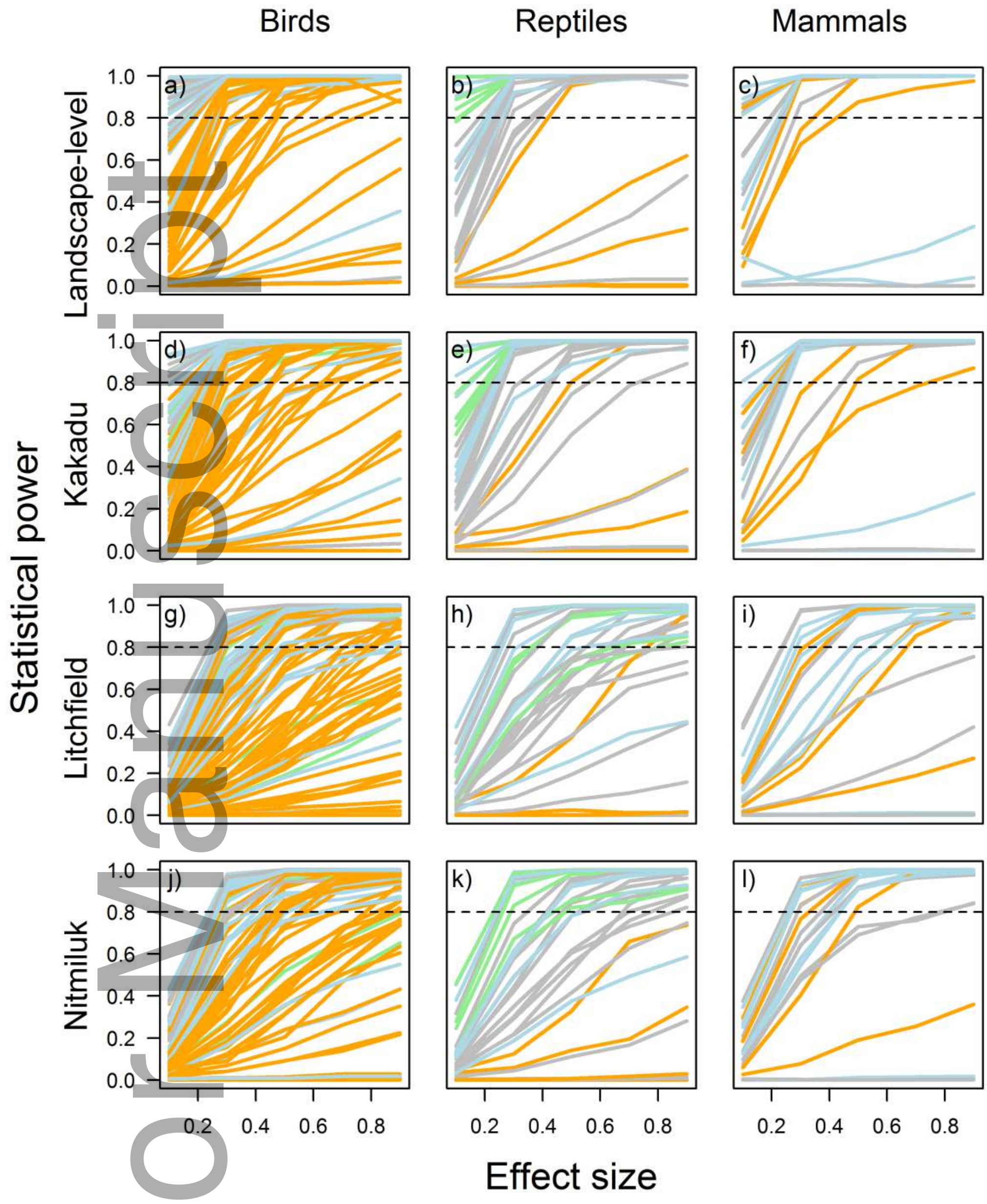
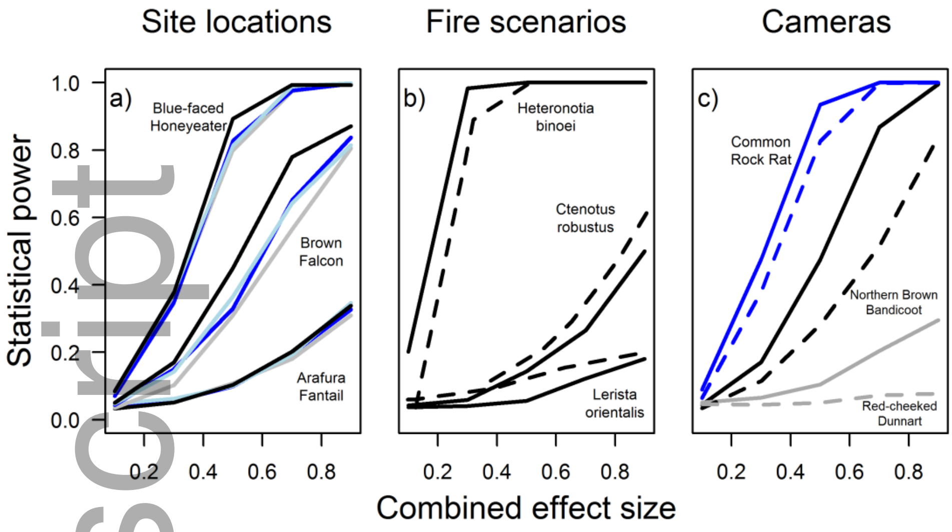

Birds
Reptiles
Mammals
Landscape-level
Kakadu
Litchfield
Nitmiluk
Statistical power
Effect size
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8

Figure 5


Site locations
Fire scenarios
Cameras
Statistical power
Combined effect size
a)
b)
c)
Blue-faced Honeyeater
Brown Falcon
Arafura Fantail
Heteronotia binoei
Ctenotus robustus
Lerista orientalis
Common Rock Rat
Northern Brown Bandicoot
Red-cheeked Dunnart
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8

698