Using Balanced Acceptance Sampling as a master sample for environmental surveys

Paul van Dam-Bates[1,*], Oliver Gansell[2], and Blair Robertson[3]

[1]*Planning Monitoring and Reporting, Department of Conservation, Christchurch, New Zealand*

[2]*Planning Monitoring and Reporting, Department of Conservation, Hamilton, New Zealand*

[3]*School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand*

[*]*Corresponding author: Paul van Dam-Bates, paul.vandambates@gmail.com*

November 10, 2017

Abstract

Well designed environmental monitoring programmes for management organisations are important for evidence based decision making. However, many environmental problems are not single agency, single spatial scale issues. A master sample can be used to coordinate and scale monitoring designs to ensure consistency in information gathered and robustness of estimators at the different spatial scales. We propose using Balanced Acceptance Sampling (BAS) to generate a master sample. Using BAS as a master sample is flexible, effective and improves on other methods previously explored. Some practical aspects of the design are addressed such as inclusion of legacy monitoring programmes, stratification, unequal probability sampling, rotating panel designs, and regional intensification. An example master sample is presented for environmental monitoring in New Zealand.

Keywords: Master Sample, Spatial Balance, Environmental Monitoring, Legacy Monitoring.

1 Introduction

Environmental management agencies rely on the results of monitoring to answer questions about the success of their policies and programmes. Monitoring is often designed to address information needs for a particular site or small set of sites. If the study is poorly designed it can result in failure to provide meaningful data to inform management and policy decision making (Legg & Nagy 2006; Nichols & Williams 2006; Field *et al.* 2007). Extrapolating from these studies to answer larger scale questions can bias the estimates as single sites are rarely representative of a broader region (Dixon *et al.* 1998; Peterson *et al.* 1999).

Increasingly, monitoring on a large scale is needed to inform management needs and assess progress towards targets concerned with changes in biodiversity globally (Noss 1999; Buckland *et al.* 2005; Pereira & Cooper 2006; Magurran *et al.* 2010). The monitoring objectives and the sample area between the national, regional, and local agencies often overlap creating efficiencies if the different groups coordinate their effort. Coordinating requires consistent formulation of goals and objectives, selection of indicators and measures, field protocols and sample design (Larsen *et al.* 2008; Fancy *et al.* 2009; Reynolds *et al.* 2016). If one agency establishes monitoring locations using standard methods and sample design, another agency can use that data for their own purposes, reducing the need to establish more monitoring. By agencies working together and through a well set out design process, as described in Reynolds *et al.* (2016), the chances of monitoring being successful are higher and concerns about extrapolating estimates from disparate data sources is also reduced.

One way to coordinate sample design is to develop a master sample; a set of points that can be sub-sampled for different monitoring activities. This was first proposed by (King 1945), but only recently has been introduced to environmental monitoring (Larsen *et al.* 2008; Theobald 2016) with implementation in the Pacific Northwest of the United States. Having different studies draw samples from the master sample has the benefit of enhancing collaboration within and between agencies to reduce duplication of effort. Additionally, consistent sample design has benefits when making estimates using data from multiple sources. Similar to providing standard field methods, the master sample provides standardised locations for sampling that ensures objective, unbiased estimation of the population parameters of interest. This coordination reduces convenience or judgement sampling and requires the user to define the objectives and sample frame clearly before gaining access to the sample points.

The sampling method chosen should be flexible enough for a variety of users and study designs to be effective for coordination. Monitoring can take place on different spatial scales such as a national monitoring programme or a local one, investigating the impact of management action. When designing an individual study, identifying heterogeneity and using stratification (Yoccoz *et al.* 2001) or unequal probability sampling (Stevens 1997) can produce more precise estimates. The study may need a unique balance of status and trend estimation which can be done by defining panels that have different revisit schema (Skalski 1990; Stevens & Olsen 1999; McDonald 2003). In all these cases the sub-samples used should be unbiased and representative.

There are many ways to generate effective samples which could be used to coordinate monitoring. A simple random sample is unbiased but is less efficient than spatially balanced designs in the presence of spatial autocorrelation (Cochran 1946; Grafström & Lundström 2013). A design is spatially balanced if the sample is well-spread over the population — a sample with few clumps and voids. A systematic sample can be considered near perfect spatial balance but is less flexible to changes in sample size making it a poor choice. Stevens & Olsen (2004) introduced Generalised Random Tessellation Stratified (GRTS) design, a spatially balanced design that is frequently used in environmental monitoring. GRTS hierarchically orders a population using a base four numbering scheme and then selects a systematic sample from the ordered population. There is also the Local Pivotal Method (LPM) (Grafström *et al.* 2012). LPM iteratively updates each sampling unit's inclusion probability in a way that makes it very unlikely to include neighbouring units in a sample. Once n units have an inclusion probability of one, the sample is released. Although the spatial balance of LPM is better than GRTS, it is computationally prohibitive on large populations. For large populations, Grafström *et al.* (2014) introduced a new rapid implementation of LPM, called suboptimal LPM. LPM has better spatial balance, but suboptimal LPM is computationally feasible on large populations. Another spatially balanced design is Balanced Acceptance Sampling (BAS) (Robertson *et al.* 2013). It uses a quasi-random number sequence to generate spatially balanced points. Similar to GRTS, the outcome of the sequence is an ordered set of points such that any contiguous sub-sample maintains spatial balance.

GRTS has been used to generate environmental monitoring master samples (Larsen *et al.* 2008). The design is particularly useful for generating master samples because GRTS points are ordered using a reverse hierarchical ordering strategy that ensures that all contiguous sub-samples are also spatially balanced (Stevens & Olsen 2004). By taking a large GRTS oversample, an ordered master sample can be obtained from which spatially balanced sub-samples can be drawn. However, once an oversample is chosen, it is not possible to generate additional points and this needs to be accounted for at the planning stage. Theobald (2016) also uses an adaptation of GRTS, Reversed Randomized Quadrant-Recursive Raster (RRQRR), implemented in ArcGIS software (Theobald *et al.* 2007) to coordinate monitoring effort. The authors' are not aware of an ordering strategy for the LPM methods and hence, it is not clear how these methods could be used for oversampling. BAS, similar to GRTS, creates an ordered set of points such that any contiguous sub-sample maintains spatial balance. To generate a master sample with BAS, a random-start is chosen and after that an infinite set of points exist for the sample. Hence, the oversample size does not need to be specified.

The purpose of this paper is to develop a master sample for environmental monitoring with a focus on terrestrial sampling of an area frame in which all sub-samples have positive area. We investigate using BAS to generate a master sample; how the points will be generated and then used in a wide variety of ways. These include adapting to different spatial scales, stratification and unequal probability sampling, changes in boundaries or resources, revisitation structure (panel design), and how to include legacy monitoring programmes. We then provide an example for how this could be applied to coordinate biodiversity monitoring at the regional and national level in New Zealand.

2 Methods

2.1 Point selection

Two-dimensional BAS points are drawn from a random-start Halton sequence $\{\mathbf{x}_k\}_{k=1}^{\infty} \subset [0,1)^2$. The ith coordinate of each point in the sequence has an associated base b_i, with $b_1 = 2$ and $b_2 = 3$. The ith coordinate of the kth point in this sequence is (Robertson et al. 2017)

$$x_k^{(i)} = \sum_{j=0}^{\infty} \left\{ \left\lfloor \frac{u_i + k}{b_i^j} \right\rfloor \mod b_i \right\} \frac{1}{b_i^{j+1}},$$

where u_i is a random non-negative integer and $\lfloor x \rfloor$ is the floor function — the largest integer that is less than or equal to x. For example, the first coordinate of the second point with $u_i = 1$ and $b_i = 2$ is

$$
\begin{aligned}
x_2^{(1)} &= \left(\left\lfloor \frac{3}{1} \right\rfloor \mod 2\right)\frac{1}{2} + \left(\left\lfloor \frac{3}{2} \right\rfloor \mod 2\right)\frac{1}{4} + \left(\left\lfloor \frac{3}{4} \right\rfloor \mod 2\right)\frac{1}{8} \\
&= \frac{1}{2} + \frac{1}{4} + 0 \\
&= \frac{3}{4}.
\end{aligned}
$$

The two-dimensional random-start Halton sequence is

$$
\{\mathbf{x}_k\}_{k=1}^{\infty} = \left\{x_k^{(1)}, x_k^{(2)}\right\}_{k=1}^{\infty}. \tag{1}
$$

Setting $u_1 = u_2 = 0$ gives the classical two-dimensional Halton sequence (Halton 1960).

The points from eqn 1 are scaled to a minimal bounding box enclosing the study area and the first n scaled points in the study area define the BAS sample. The BAS points are kept in the same order as they appear in eqn 1 and will have good spatial spread over the study area. Furthermore, any continuous subset of the BAS sample will also have good spatial spread (Robertson *et al.* 2017).

The random integer vector in the sequence $\mathbf{u} = (u_1, u_2) \in [0, 10^7]^2$ is chosen so that $\mathbf{x}_1$ falls within the sample area (Robertson *et al.* 2017). This gives $\lambda 10^{14}$ possible BAS samples of size n, where λ is the fraction of the bounding box occupied by the study area. By ensuring the random start comes from a large set of integers, the BAS points are uniformly distributed (Robertson *et al.* 2013). Once the random-start is selected an infinite number of BAS points exist over the study which constitutes the master sample. Higher dimensional points can be defined by using different co-prime bases for each additional dimension (e.g. $b_3 = 5$ when sampling from a $[0, 1)^3$).

2.2 Spatial Scales

The master sample should work at different spatial scales to answer national, regional, and local objectives. Let A be a measurable subset of the study area for which the master sample is defined. Because the master sample $\{\mathbf{x}_k\}_{k=1}^{\infty}$ is uniformly distributed over $[0,1)^2$ (Wang & Hickernell 2000) and A is measurable, there exists a subsequence $\{\mathbf{z}_j\}_{j=1}^{\infty} \subset \{\mathbf{x}_k\}_{k=1}^{\infty}$ such that each $\mathbf{z}_j \in A$. Furthermore, $\{\mathbf{z}_j\}_{j=1}^{n}$ is a BAS sample of size n drawn from A, with its random start and bounding box defined by the master sample. Hence, BAS samples can be drawn from the master sample at any spatial scale within the study area of the master sample. This also means that a national sample can share points with monitoring at the local level (see Section 2.4).

2.3 Stratification and Unequal probability

Stratification with the master sample is essentially the same as taking a subsample for a specific measurable subset of the study area as described above. The ith stratum (measurable) has a subsequence $\{\mathbf{z}_j\}_{j=1}^{\infty} \subset \{\mathbf{x}_k\}_{k=1}^{\infty}$ such that each $\mathbf{z}_j$ is in the stratum. The BAS sample for the ith stratum is $\{\mathbf{z}_j\}_{j=1}^{n_i}$, where n_i is the sample size required. Hence, each stratum has its own BAS sample with its random start and bounding box defined by the master sample.

If unequal probability sampling is required, a third dimension is added to the bounding box. This extra dimension allows BAS to sample from an arbitrary inclusion density function $\pi(\mathbf{x})$ using an acceptance/rejection sampling strategy (Robertson *et al.* 2013). Specifically, a point $\mathbf{x}_k = (x_k^{(1)}, x_k^{(2)}, x_k^{(3)})$ is accepted if $(x_k^{(1)}, x_k^{(2)})$ is in the study area and $\pi(x_k^{(1)}, x_k^{(2)}) \leq \alpha x_k^{(3)}$, where α is a scaling factor to ensure $\max_{\mathbf{x}} \pi(\mathbf{x}) = 1$. The impact of this is that some of the master sample points in eqn(1 will be skipped. Skipping points in eqn 1 changes the distribution of points in each BAS sample, with fewer points being drawn from areas where $\pi(\mathbf{x})$ is low.

2.4 Changing boundaries and resources

For long-term monitoring programmes, the boundaries of study regions may change over time. This is easy to accommodate with the master sample, provided the changes are within the initial bounding box. Let A be a measurable study area whose boundaries changed, defining a new measurable study area B with $A \cap B \neq \emptyset$. If there are no sampled BAS points in $A \cap B$, points from the master sample are drawn to sample B. Otherwise, let $\mathbf{x}_k$ be the sampled point in $A \cap B$ with the largest index k. A BAS sample in B, that includes sampled points from $A \cap B$, is achieved if all the master sample points that fall in B with indices less than k are sampled. If a smaller sample is desired in B, potentially due to a change in resources, then points with the larger indices in B are removed (See Figure 1). If more points are required, then points can be added from the master sample in B until the new sample size is achieved. Ensuring that BAS samples are drawn from each study region means that spatial balance and good sampling properties are maintained.

2.5 Panel design

In environmental surveys that are repeated through time, some samples may be visited more frequently than others. Estimates of status and/or trend can be improved by balancing the number of new points sampled each year with repeated sampling on existing points (Urquhart & Kincaid 1999). A panel is defined as all samples that have the same visitation schedule. The points within a panel as well as between panels should be representative and unbiased. A panel design is achieved using the master sample by choosing the subset of points $\{\mathbf{z}_j\}_{j=1}^{\infty}$ that fall within the the study area and the order of the points define each panel. For panel 1 we use $\{\mathbf{z}_j\}_{j=1}^{n_1}$, for panel 2 we use $\{\mathbf{z}_j\}_{j=1+n_1}^{n_1+n_2}$ and so on, where n_i is the sample size for the ith panel. By defining the panels this way the overall sample is still a BAS design as well each panel. Once a full rotation of all samples has been carried out, if additional points are needed, they are added from the unsampled points in the master sample in the order that they appear. Note, when this occurs each panel may not be a true BAS sample but as shown in the legacy units simulation below, adding BAS points to an existing sample does not significantly impact estimation and the sample will still be equi-probable and give unbiased estimators. If budgets change, points should be removed by last in, first out. Table 1 shows an example panel design.

2.6 Incorporating legacy monitoring

A master sample is intended for coordinating large scale monitoring. Often, there is legacy monitoring that may already be well designed and this should be accommodated. We will consider two different types of legacy monitoring designs: simple random sampling (SRS) and random-start systematic sampling (SS). These are equi-probable designs, where each sampling unit has an equal chance of being included in a sample. If the existing monitoring is insufficient, the BAS master sample can be used to draw additional units from the area. Furthermore, the inclusion probabilities around the legacy units can be decreased to reduce the chance of drawing units close to the legacy units (Foster *et al.* 2017). Using the unequal probability sampling method described above, we can incorporate this method into the master sample. We will call the altered inclusion probability sampling method aBAS.

For unbiased estimation, inclusion probabilities need to be specified. Using an existing n_l legacy units and augmenting with n_b BAS units from a population of N units, we assume the legacy units have equal inclusion probabilities $\pi_l = (n_l + n_b)/N$. The inclusion probabilities for the remaining units are then specified by the user so $\sum_{i=1}^{N} \pi_i = n_l + n_b = n$. For example, an equi-probable design uses $\pi_i = (n_l + n_b)/N$ for $i = 1, 2, ..., N$. By making an assumption about the legacy monitoring and generating BAS points independently, the design and analysis is simple. However, incorporating legacy units this way may decrease the overall spatial balance of the sample which in the presence of strong spatial autocorrelation can impact precision and is why aBAS has better precision (Foster *et al.* 2017).

Spatially balanced designs use a local neighbourhood variance (LNV) (Stevens & Olsen 2003) and as the proportion of legacy units is increased the LNV underestimates true variance (Foster *et al.* 2017). If the legacy units and the augmented samples are considered separately with sample size n_l and n_b ($n_l + n_b = n$) respectively, then the mean can be written as

$$\hat{y} = \frac{n_l}{n} \sum_{i=1}^{n_l} \frac{y_l}{N\pi^*_{li}} + \frac{n_b}{n} \sum_{i=1}^{n_b} \frac{y_b}{N\pi^*_{bi}} \tag{2}$$

where the inclusion probabilities are rescaled $\pi^*_{li} = \pi_i \times \frac{n}{n_l}$ and $\pi^*_{bi} = \pi_i \times \frac{n}{n_b}$. It then follows that the variance can be estimated as

$$\widehat{\mathrm{var}}(\hat{y}) = \left(\frac{n_l}{n}\right)^2 \times \widehat{\mathrm{var}}(\hat{y}_l) + \left(\frac{n_b}{n}\right)^2 \times \widehat{\mathrm{var}}(\hat{y}_b). \tag{3}$$

Here $\widehat{\mathrm{var}}(\hat{y}_l)$ is the Sen, Yates and Grundy variance estimate using the legacy units and $\widehat{\mathrm{var}}(\hat{y}_b)$ is the LNV variance estimate using spatially balanced units.

A simulation study was carried out to investigate the different methods to incorporate legacy monitoring. The sampling frame was defined as a 100×100 raster in $[0, 1)^2$ and an estimate of the population mean and standard error was sought. The response value for each raster cell was defined as the integral of $f(\boldsymbol{x})$ over the cell. Three different functions were considered, a strong spatial trend (Robertson *et al.* 2013; Grafström *et al.* 2012), the Peak function, and the Bird function. These functions are given in the online supplementary material section. Scenarios similar to Foster *et al.* (2017) using the program R (R Core Team 2015) were run. We used an overall sample size of $n = 60$. A number of legacy units ($n_l \in 3, 4, ..., 57$) were generated either as SRS or SS. More units ($n_b = 60 - n_l$) were generated using GRTS (Kincaid & Olsen 2016), BAS, aBAS (Foster 2016) and SRS to achieve the full sample size of $n = 60$. Each scenario was run 1000 times. When estimating the standard error of SS units we assumed a simple random sample, which provides conservative estimates in the presence of spatial autocorrelation (Aune-Lundberg & Strand 2014; Strand 2017). A detailed description of the simulation and functions used can be accessed in the supplementary material.

The results of the simulation can be seen in Figure 2. In all cases, adding some spread to the points improved precision over using SRS. aBAS performed the best when SRS legacy units were used, but had similar performances to BAS and GRTS for SS. The SS legacy units already had good spread and aBAS was not necessary to force better overall spatial balance. Population 3 has periodic structure, which made systematic sampling perform poorly because the spread of the legacy units matched the structure of the population. The standard error estimate from eqn 3 was relatively accurate for SRS, although conservative for aBAS and overestimated as expected for SS. The legacy monitoring was improved by augmenting the samples with spatially balanced points. Given that the legacy monitoring is well designed, there are no issues with incorporating legacy samples with the master sample.

We recommend using aBAS for single measurements or simple repeated studies when legacy monitoring is SRS or poorly spread. However, it is important that inclusion probabilities are positively correlated to the response variable (Robertson *et al.* 2017) and this is not the case for ordered subsets of aBAS. As a result, a rotating panel design may lose precision using aBAS when legacy units are not sampled on each occasion. For estimation with aBAS we assumed equal inclusion probabilities which is only met when the sample is made up of a single equi-probable random legacy sample followed by a single aBAS sample. Multiple samples of aBAS on the same legacy units would violate this as well as any issues in the process for how the legacy monitoring was designed. In general, we recommend to generate spatially balanced points independently of the legacy monitoring and combine the two methods only when the legacy monitoring is well designed. For design-based estimation, the variance estimate in eqn 3 should be used. This corrects the underestimation of LNV but requires there are four or more spatially balanced points as suggested in the spsurvey package in R (Kincaid & Olsen 2016) and three or more legacy points. Otherwise, use LNV ($n_l < 3$) or the legacy variance estimator ($n_b < 4$) for all points. In practice, model-based estimation is often used and any analyses should account for the spatial aspect of the design as in Foster *et al.* (2017).

3 Application: New Zealand terrestrial monitoring

The New Zealand Department of Conservation (DOC) is the lead biodiversity management agency in New Zealand (NZ), responsible for managing $\approx 30\%$ of New Zealand as public conservation land (PCL). Development of a national monitoring system has exposed the challenges in coordinating monitoring design to provide results meaningful at a local, regional and national scale. Increasingly partner environmental agencies (local government etc.) and central government expect cross-agency collaboration and coordination of systems and processes. Considerable effort has gone into a coordinated approach for indicators and measures and field protocols (Department of Conservation 2016). There is an existing national sample of PCL, known as the National Level Monitoring (NLM) programme which is an 8-km systematic grid of $\approx$ 1400 monitoring sites (Coomes *et al.* 2002). The NLM programme focusses on status and trend monitoring at the national scale of key indicators of ecosystem structure and composition. Through ongoing collaboration with local government agencies, the grid has been extended off PCL.

DOC ecosystem and species management is focussed on a suite of priority sites known as Ecosystem Management Units (EMUs). To assess the outcome of management interventions across all EMUs, the 8-km grid NLM needs to be intensified with changes to visitation frequency and methods relating to the specific management monitoring objectives. At the same time, there are EMU specific questions to be answered that require another intensification to a regional/local level. For example, one of DOC's EMUs, Abel Tasman National Park (ATNP) (40°56'03.8"S 172°58'19.7"E), is managed in partnership with a philanthropic foundation. The agreement governing this partnership requires the foundation to invest in the recovery of biodiversity in ATNP. Once it is shown that the targets pertaining to increases in the abundance and distribution of bird species and vegetation condition are met, DOC will be responsible for the maintenance of biodiversity in ATNP. Monitoring is required to establish the current state of key indicators in ATNP and then determine when agreed targets have been achieved. A sample design is required which can incorporate the existing NLM monitoring, sample broad scales and enable local intensification of monitoring.

A national sample of New Zealand using BAS can make up the master sample, with the existing NLM points included. For efficiency and simplicity, each island (North Island, South Island, Stewart Island, Chatham Islands, etc.) will be stratified and have their own bounding box and random seed. The seed chosen for the South Island was $\mathbf{u} = (4887260, 18041662)$ with minimum bounding box in New Zealand Transverse Mercator 2000 (NZTM2000)

$$[1089354, 1721164] \times [4747979, 5516919].$$

These values define the scaling needed to map the random start Halton points to the South Island of New Zealand (NZ). For example, the first point is

$$\begin{aligned}\mathbf{x}_1 &= (631810x_1^{(1)} + 1089354, 768940x_1^{(2)} + 4747979) \\ &\approx (1235673, 5075613).\end{aligned}$$

To sample EMUs in NZ using the master sample, select all sites that fall within EMU polygons. The actual required sample size and visitation frequency should reflect the monitoring objectives and follow a similar process as outlined by (Reynolds *et al.* 2016). See Figure 3 for an example of clipping the master sample on the South Island into the first 500 samples that fall within EMUs.

At ATNP, one of the key targets is focussed on bird abundance and distribution through the park. A sample size of $n = 65$ was chosen based on a precision analysis using simr (Green & MacLeod 2016) and historical bird count data from an existing intensively monitored site which showed that temporal variation was less than spatial. Therefore, 15 samples were selected to be measured annually and the other 50 on a rotating 5-year panel $[1-0,(1-4)^5]$. From the NLM programme mentioned above, there were four legacy samples in the sample frame which would be included in the rotating panel on years corresponding to the years they are to be sampled. If DOC implemented an EMU monitoring programme of 500 augmented points on the South Island, then there would be an additional seven points in ATNP monitored and funded by DOC using the master sample. The EMU points would make up Panel 1 based on the hierarchical order of the master sample. See Figure 4 for the selected points in this monitoring programme. ATNP is an example of localised monitoring using the master sample that can contribute to national estimates of bird abundance and distribution. DOC gets better precision with the increased sampling in ATNP and the philanthropic foundation saves resources by using DOC's national investment in monitoring explicitly in their design.

The master sample above is entirely defined by the seed $\mathbf{u}$ and the bounding box. Hence, there is no need for a repository to hold the coordinates. Computationally the master sample is easy to run on the fly. Generating 65 points for ATNP in Figure 4 takes ≈ 0.5 seconds on a desktop computer. See supplementary materials for R script to generate a master sample in New Zealand.

4 Discussion

A master sample can be a useful tool to organise environmental monitoring at different spatial scales as previously done using GRTS or RRQRR (Larsen *et al.* 2008; Theobald 2016). By using BAS instead of GRTS there is better spatial balance (Robertson *et al.* 2013) and no need for an oversample. It is also possible to add an extra dimension to do unequal probability sampling leading to an overall more flexible design. Depending on the population being sampled, spatial balance may or may not lead to a more precise sample than simple random sampling. However, it is likely that some spatial balance will improve efficiency for most designs. Not needing to oversample to create a master sample using BAS means that it will remain relevant at any scale that monitoring takes place no matter how localised. BAS can be used for sampling three-dimensional space (Robertson *et al.* 2013) which generalises the concepts presented here to work for atmospheric or oceanic monitoring.

Previous master samples rely on large source files for point coordinates. BAS does not because it is deterministic once the random seed is chosen. Generating a BAS master sample in R is computationally quick and easy to program making it possible for a user to run a function in R (see supplementary online material section) to sample a chosen region from a shape file. Note, by making use of the deterministic nature of the Halton sequence and Halton boxes (Robertson *et al.* 2017) the code can be made computationally efficient for any set of shape files and sample sizes required.

In our experience, any large scale long-term monitoring will need to incorporate already existing monitoring programmes that are proven effective. This was a requirement in developing a master sample for New Zealand. We have shown that there is no major issue with incorporating legacy monitoring into the design but recommend that all sites are rigorously vetted to ensure no known biases are included if historically the sites were judgement samples or for other reasons. Using panel designs can help incorporate the already existing visitation schedule of the legacy sites into an efficient monitoring design. By generating the master sample independently of the legacy monitoring it is possible that a legacy site and master sample site could be close, in this case both sites still need to be sampled.

The master sample helps coordinate the points sampled for environmental surveys. Every survey at the local and national level should still go through rigorous design. This means defining the objectives of monitoring clearly and the methods to use so that they are consistent with standard methodology as required by the objective. By following the steps outlined in Reynolds *et al.* (2016) and using the master sample for point generation, we believe that the monitoring programmes undertaken at all levels will have improved efficiency and contribute to the overall knowledge of the population of interest.

5 Acknowledgements

This research was funded by the Department of Conservation in New Zealand. We would like to thank Elaine Wright for her support, Tony Olsen, Scott Foster and Amy Hawcroft for their advice.

References

Aune-Lundberg, L. & Strand, G.H. (2014) Comparison of variance estimation methods for use with two-dimensional systematic sampling of land use/land cover data. *Environmental modelling & software*, **61**, 87–97.

Buckland, S., Magurran, A., Green, R. & Fewster, R. (2005) Monitoring change in biodiversity through composite indices. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **360**, 243–254.

Cochran, W.G. (1946) Relative accuracy of systematic and stratified random samples for a certain class of populations. *The Annals of Mathematical Statistics*, pp. 164–177.

Coomes, D.A., Allen, R.B., Scott, N.A., Goulding, C. & Beets, P. (2002) Designing systems to monitor carbon stocks in forests and shrublands. *Forest Ecology and Management*, **164**, 89–108.

Department of Conservation (2016) Monitoring species and habitats in terrestrial, freshwater and marine environments. `http://www.doc.govt.nz/our-work/biodiversity-inventory-and-monitoring/`. Accessed: 2016-09-29.

Dixon, P.M., Olsen, A.R. & Kahn, B.M. (1998) Measuring trends in ecological resources. *Ecological Applications*, pp. 225–227.

Fancy, S., Gross, J. & Carter, S. (2009) Monitoring the condition of natural resources in us national parks. *Environmental Monitoring and Assessment*, **151**, 161–174.

Field, S.A., O'Connor, P.J., Tyre, A.J. & Possingham, H.P. (2007) Making monitoring meaningful. *Austral Ecology*, **32**, 485–491.

Foster, S.D. (2016) *MBHdesign: Spatial Designs for Ecological and Environmental Surveys*. R package version 1.0.61.

Foster, S.D., Hosack, G.R., Lawrence, E., Przeslawski, R., Hedge, P., Caley, M.J., Barrett, N.S., Williams, A., Li, J., Lynch, T. *et al.* (2017) Spatially-balanced designs that incorporate legacy sites. *Methods in Ecology and Evolution*.

Grafström, A., Lundström, N.L.P. & Schelin, L. (2012) Spatially balanced sampling through the pivotal method. *Biometrics*, **68**, 514–520.

Grafström, A. & Lundström, N.L. (2013) Why well spread probability samples are balanced. *Open Journal of Statistics*, **3**, 36–41.

Grafström, A., Saarela, S. & Ene, L.T. (2014) Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. *Canadian Journal of Forest Research*, **44**, 1156–1164.

Green, P. & MacLeod, C.J. (2016) simr: an r package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, **7**, 493–498.

Halton, J.H. (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, **2**, 84–90.

Kincaid, T.M. & Olsen, A.R. (2016) *spsurvey: Spatial Survey Design and Analysis*. R package version 3.3.

King, A.J. (1945) The master sample of agriculture: I. development and use. *Journal of the American Statistical Association*, **40**, 38–45.

Larsen, D.P., Olsen, A.R. & Stevens, D. (2008) Using a master sample to integrate stream monitoring programs. *Journal of Agricultural, Biological, and Environmental Statistics*, **13**, 243–254.

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.

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J. & Watt, A.D. (2010) Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in ecology & evolution*, **25**, 574–582.

McDonald, T.L. (2003) Review of environmental monitoring methods: survey designs. *Environmental monitoring and assessment*, **85**, 277–292.

Nichols, J.D. & Williams, B.K. (2006) Monitoring for conservation. *Trends in ecology & evolution*, **21**, 668–673.

Noss, R.F. (1999) Assessing and monitoring forest biodiversity: a suggested framework and indicators. *Forest ecology and management*, **115**, 135–146.

Pereira, H.M. & Cooper, H.D. (2006) Towards the global monitoring of biodiversity change. *Trends in Ecology & Evolution*, **21**, 123–129.

Peterson, S.A., Urquhart, N.S. & Welch, E.B. (1999) Sample representativeness: a must for reliable regional lake condition estimates. *Environmental science & technology*, **33**, 1559–1565.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reynolds, J.H., Knutson, M.G., Newman, K.B., Silverman, E.D. & Thompson, W.L. (2016) A road map for designing and implementing a biological monitoring program. *Environmental Monitoring and Assessment*, **188**, 1–25.

Robertson, B.L., Brown, J.A., McDonald, T. & Jaksons, P. (2013) BAS: Balanced acceptance sampling of natural resources. *Biometrics*, **69**, 776–784.

Robertson, B., McDonald, T., Price, C. & Brown, J. (2017) A modification of balanced acceptance sampling. *Statistics & Probability Letters*.

Skalski, J.R. (1990) A design for long-term status and trends monitoring. *Journal of Environmental Management*, **30**, 139–144.

Stevens, D. (1997) Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, **8**, 167–195.

Stevens, D. & Olsen, A.R. (2003) Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, **14**, 593–610.

Stevens, D. & Olsen, A.R. (2004) Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, **99**, 262–278.

Stevens, D. & Olsen, A.R. (1999) Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, **4**, 415–428.

Strand, G.H. (2017) A study of variance estimation methods for systematic spatial sampling. *Spatial Statistics*, **21**, 226–240.

Theobald, D.M. (2016) A general-purpose spatial survey design for collaborative science and monitoring of global environmental change: The global grid. *Remote Sensing*, **8**, 813.

Theobald, D.M., Stevens, D.L., White, D., Urquhart, N.S., Olsen, A.R. & Norman, J.B. (2007) Using gis to generate spatially balanced random survey designs for natural resource applications. *Environmental Management*, **40**, 134–146.

Urquhart, N.S. & Kincaid, T.M. (1999) Designs for detecting trend from repeated surveys of ecological resources. *Journal of Agricultural, Biological, and Environmental Statistics*, pp. 404–414.

Wang, X. & Hickernell, F.J. (2000) Randomized halton sequences. *Mathematical and Computer Modelling*, **32**, 887–899.

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

6 Supporting Information

More information about the simulation study and the R code used to generate the Master Sample for New Zealand can be found in the supplementary text.

7 Data Availability

We present a maintained version of the R code to generate a master sample in New Zealand, including a shape file example on github for access to readers familiar with R. `https://github.com/ogansell/MSampNZ`.

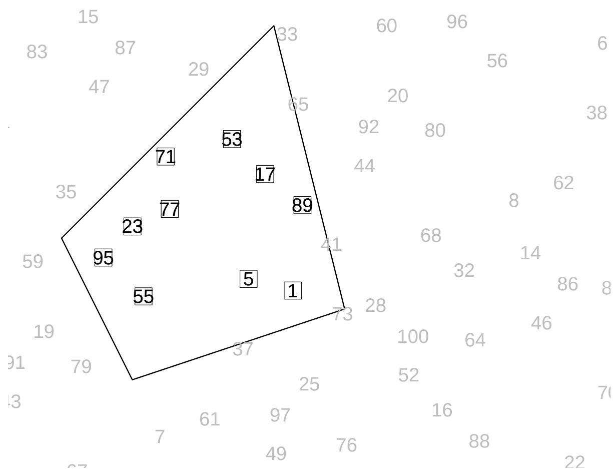
8 Author Contribution

All authors had significant contributions to this research. The paper was initiated by a requirement from the Department of Conservation (DOC) for coordinated monitoring of ecosystems at the national level. This work fell directly to Ollie Gansel and Paul van Dam-Bates who collaborated with Blair Robertson. We were funded as salary through our roles at DOC and the University of Canterbury.

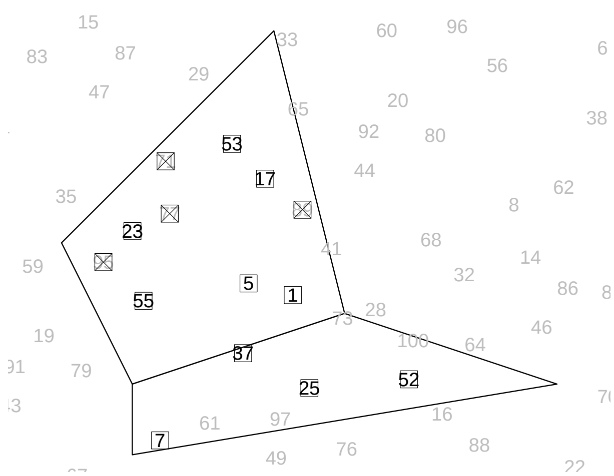
Figures and Tables

Table 1: Example of a panel design in which panel 1 is sampled annually and panels 2-4 are sampled with a 2 year break in between described as $[1-0,(1-2)^3]$ in McDonald (2003). The sample size n and the points from the master sample in the study area $\{\mathbf{z_j}\}$ are shown below, where an X indicates that the panel is sampled on that occasion.

			Sample Occasion									
Panel	n	Sample	1	2	3	4	5	6	7	8	9	10
1	20	$\{\mathbf{z_j}\}_{j=1}^{20}$	X	X	X	X	X	X	X	X	X	X
2	10	$\{\mathbf{z_j}\}_{j=21}^{30}$	X			X			X			X
3	10	$\{\mathbf{z_j}\}_{j=31}^{40}$		X			X			X		
4	10	$\{\mathbf{z_j}\}_{j=41}^{50}$			X			X			X	



(a)



(b)

Figure 1: An example of changing boundaries using a BAS the master sample where each number denotes the index of the BAS point in the master sample. (a) shows 10 points drawn from the master sample in the study region. (b) shows an enlarged region and the 10 points drawn from the master sample in the region. Points 71,77,89 and 95 are removed and 7,25,37 and 52 are included in the new region. If the sample size is increased to 14, then points 61,71,77 and 89 would be included.

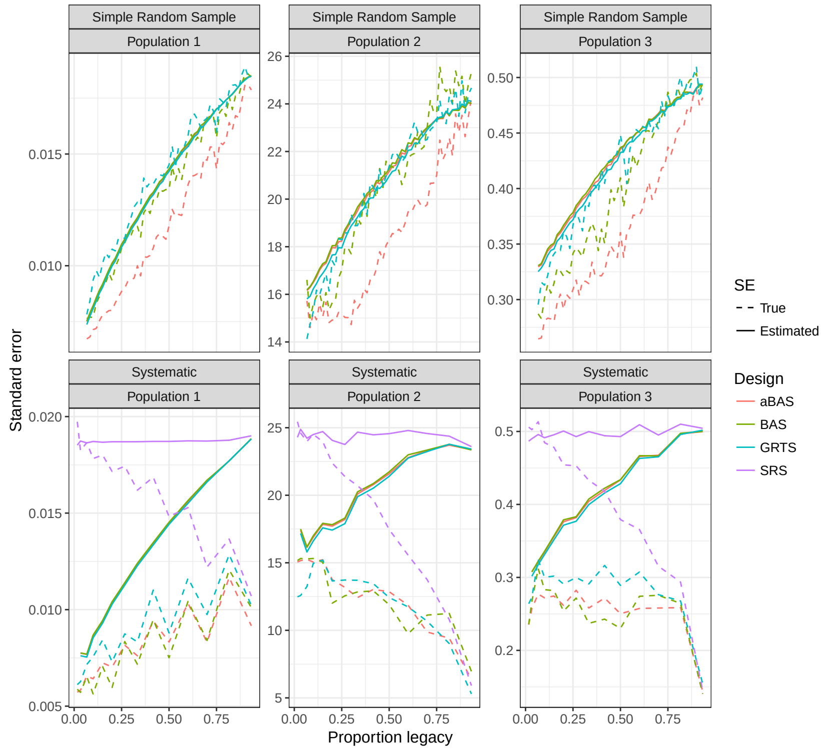
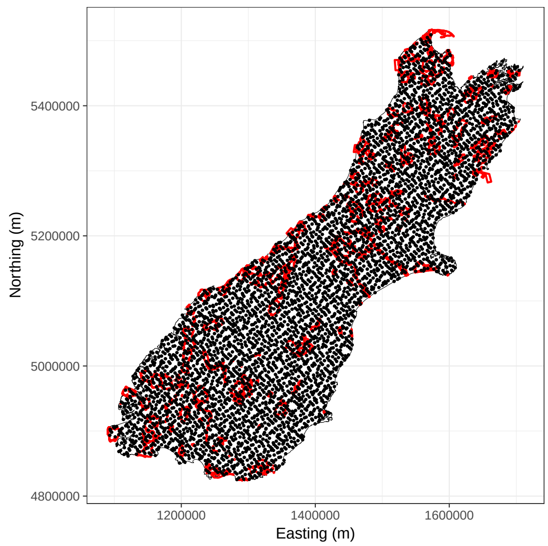
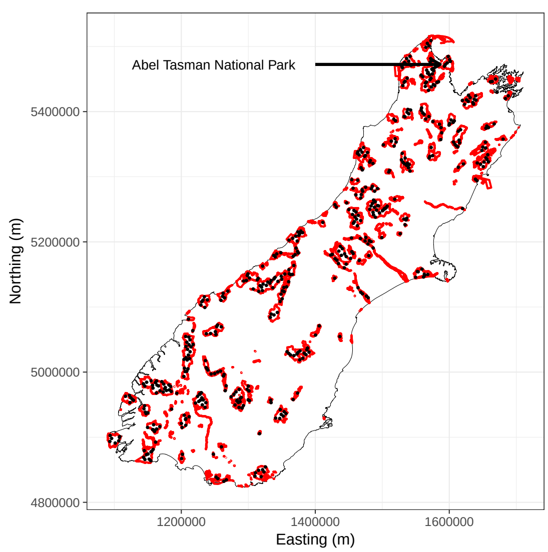


Figure 2: Results from the simulation study testing the impact of adding new samples from altered Balanced Acceptance Sampling (aBAS), Balanced Acceptance Sampling (BAS), Generalised Random Tessellation Stratified (GRTS), and simple random sampling (SRS) to existing legacy monitoring units. Three populations with varying spatial structure were tested. Population 1, a strong spatial trend. Population 2, a peak function. Population 3, a cyclical (bird) function. The estimated and simulated (True) standard errors are shown.



(a)



(b)

Figure 3: South Island of New Zealand. (a) shows the first 5000 points of the master sample overlayed on red ecosystem management units (EMUs). (b) shows 500 master sample points from (a) that fall within the EMUs in red. Abel Tasman National Park receives seven points which are included as panel 1 points in Figure 4

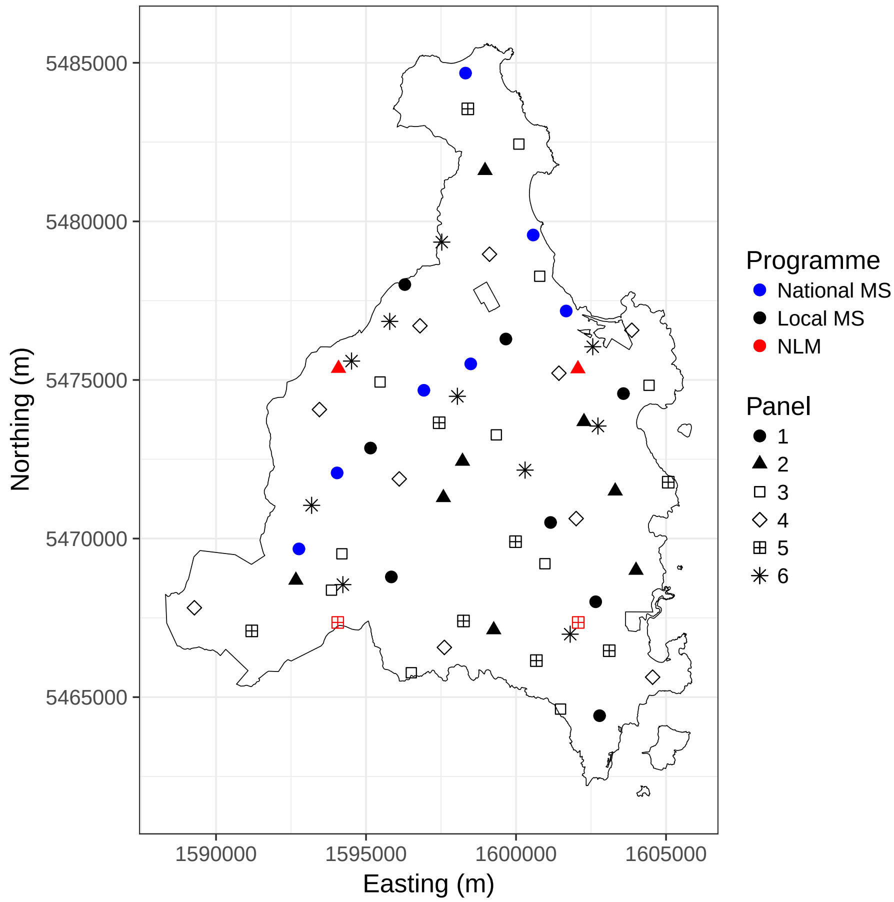


Figure 4: An example of bird monitoring in Abel Tasman National Park New Zealand. Panel 1 is measured annually while the other panels are on a 5-year rotation described as $[1-0,(1-4)^5]$. In the first year, panels 1 and 2 would be measured. Blue points are master sample points measured by a national Ecosystem Management Unit monitoring programme, see Figure 3. The red points are from the 8-km systematic grid National Level Monitoring programme. This design gives excellent spatial coverage over the park each year ($n = 25$) and over a 5-year period ($n = 65$).