Using balanced acceptance sampling as a master sample to coordinate New Zealand environmental monitoring programmes

Paul van Dam-Bates, Oliver Gansell, & Blair Robertson

April 14, 2017

Abstract

Environmental monitoring for management organisations like the Department of Conservation is critical. Without good information about outcomes, poor management actions may persist much longer than they should or initial intervention may occur too late. The Department currently conducts focused research at key natural heritage sites (Tier 3) as well as a long term national monitoring (Tier 1). The link between the two tiers of investigation to assess the impact of management across New Zealand (Tier 2) is yet to be implemented but faces unique challenges for working at many different spatial scales and coordinating with multiple agencies. The solution is to implement a Master Sample using Balanced Acceptance Sampling (BAS). To do this some practical aspects of the sample design are addressed such as stratification, unequal probability sampling, rotating panel designs and regional intensification. Incorporating information from Tier 1 monitoring directly is also discussed.

1 Introduction

Conservation management groups rely on the results of environmental monitoring to answer questions about the success of their programs. Often this monitoring is not properly designed and fails to provide meaningful information to management [9, 11, 3]. Extrapolating from these studies to answer larger scale questions can introduce bias into the estimates [14]. The material exists to overcome these problems when designing a monitoring programme [4, 16, 23] but that doesn't ensure the advice will be used. Providing specific tools for standard problems can help such as the Department of Conservation's (DOC) monitoring toolbox [12]. By coordinating monitoring effort between agencies and guiding them through the design process the chances of success are higher. This can also lead to planned extrapolation of data allowing for robust estimates at larger scales borrowing from the information gained at the regional level.

1

One way to coordinate monitoring effort is to develop a Master Sample; a set of points that can be subsampled for different monitoring activities. This idea has been around for a while [7] but only recently has been introduced to environmental monitoring [8] with implementation in the Pacific Northwest of the United States. Having different studies draw samples from the Master Sample has the benefit of collaborating between groups to reduce effort as well as consistent methodology when extrapolating from multiple data sources. Choosing the sampling procedure for the Master Sample is critical for success. A good sample design allows for unbiased estimation of the population of interest. Simple random sampling achieves this but is less efficient than designs that are well spread across the population in the presence of spatial autocorrelation.

Generalized random tessellation stratified design (GRTS) [21] provides excellent spatial balance and existing environmental monitoring Master Samples [8] use it. Points produced by GRTS maintain spatial balance for any ordered subset due to reversed hierarchical order [21]. By taking a large oversample on a GRTS draw, a Master Sample can be obtained. Once an oversample is chosen, it is not possible to generate points with a higher density which needs to be accounted for at the planning stage. Another spatially balanced design is Balanced Acceptance Sampling (BAS) [17]. It uses a random-start quasi-random number sequence which creates an even spread of points across the sample space. Similar to GRTS, the outcome of the sequence is an ordered set of points such that any subset maintains spatial balance but must be sampled in order. To implement a Master Sample with BAS, a random-start is chosen and after that an infinite set of points exist for the sample.

The sampling method chosen must be flexible enough for a variety of users and study designs. Monitoring can take place on different spatial scales such as a national monitoring programme or a regional one researching a local management action. When designing an individual study, identifying heterogeneity and using stratification [24] or unequal probability sampling [20] 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 [19, 10, 22]. In all these cases the subsamples used must unbiased and representative.

The purpose of this paper is to develop a Master Sample for environmental monitoring for management outcomes in New Zealand. The Department of Conservation is tasked with managing New Zealand's natural heritage and is responsible for Public Conservation Land (PCL) which makes up approximately one-third of all land. However, regional councils and non-government organisations like the NEXT Foundation also partake in environmental management and monitoring activity. A Master Sample across all of New Zealand will help coordinate monitoring effort between agencies and build a bigger picture story about outcomes. In this paper we will investigate BAS as a Master Sample; how the points will be generated for New Zealand as well as how they can be

used in different monitoring designs. We will then provide examples for how the Master Sample could be applied at the regional and national level.

2 Methods

2.1 Point Selection

Balanced Acceptance Sampling points are drawn from a random-start Halton sequence $\{z_k\}_{k=0}^{\infty}$ which is defined by a co-prime base chosen as $(2,3)$. By choosing this base the bounding box of the sample is divided into six equal sized rectangles into which points are chosen in a systematic order as shown in Figure 3. The random element is the start of the sequence which should be a point that falls within the sample area [18]. The sequence generates quasi-random points in $[0,1]^2$ with good spatial spread which are mapped to the bounding box of the sample area. Points that do not fall into that area are removed via acceptance/rejection sampling. Once the random-start is selected an infinite number of BAS points exist which constitutes the Master Sample. For efficiency New Zealand will be stratified by island with a random seed selected for each island (North Island, South Island, Stewart Island, Chatam Islands, etc.).

The reason the random-start must fall within the sample region is so that any realisation of BAS is an equiprobable sample. By allowing a start anywhere in the bounding box, the points that are rejected increase the probability that a particular area in the sample gets selected [18]. For example, in Figure 3, if region B is removed from the sample, every time a point is randomly selected in B and rejected, the next point in the sequence is in C which increases the probability of the first point ending up in C. This is important for simulating different cases of the Master Sample and testing the theory of BAS. However, once the BAS sample is selected, all subsamples maintain the BAS properties as long as the points remain in order. By choosing the random start from a very large set of integers, any point within the sample has an equal probability of being selected. Once the sequence start is chosen, any region has a dense set of points that fall within it with uniform inclusion densities [17].

2.2 Incorporating Pre-existing Monitoring Programs

There currently exists a national monitoring programme in New Zealand administered jointly by the Ministry for the Environment (MFE) and DOC for the purpose of carbon monitoring on an 8-km systematic grid [2]. The Department of Conservation added monitoring of birds, mammals and vegetation at 1400 points falling within the PCL with a $[(1-4)^5]$ revisit scheme where a single panel is measured each year (Tier 1 monitoring). A Master Sample for New Zealand should accommodate this existing programme to improve monitoring

efficiency. This can be done by recognising that both BAS and a random-start systematic grid create an equiprobable sample that are unbiased and representative. The BAS points can be used to intensify the pre-existing sample in order to answer more specific questions about management success. If the Tier points are selected first and then the BAS points added, inclusion probabilities π can be approximated by thinking in discrete terms with an arbitrary N discrete sample units in the population to be selected with n_T Tier 1 points and n_B BAS points written as $\pi = P(T \cup B) = \frac{n_T + n_B}{N}$.

The impact of intensifying sampling on the systematic grid using BAS was investigated by simulation. One hundred and twenty-one points on $[0, 80]^2$ were generated systematically using the package sp in the program R [15, 13, 1]. Simple Random Sampling (SRS), GRTS [6], and BAS points were added until spatial variance became stable ($n = 350$). Spatial balance was measured using the variance of the Voronoi polygons [17, 21, 5] and was repeated 300 times. Balanced Acceptance sampling was the best of the three methods for maintaining spatial balance Figure 1. An example of what the intensified grid looked like with the Voronoi polygons can be seen in Figure 2.

2.3 Spatial Scales

The Master sample should work at different spatial scales to answer national questions about the effect of management action and also at the the local level to assess a single management activity such as a pest control. At any scale where a space A is defined as the sample frame, starting from the Master Sample $\chi = \{x_k\}_{k=0}^{\infty}$ a subsequence can be chosen $\{y_j \subset \chi : y_j \in A\}$. Thus, an infinite set of points can be selected for the sample frame that is still a true BAS sample with the random-start chosen by the Master Sample and a bounding box that is no longer minimum size. The order of y_j is still important for spatial balance and the sample maintains uniform density.

2.4 Stratification and Unequal probability

Stratification with the Master Sample is essentially the same as taking a subsample for a specific sample frame as described above. Define your strata sample frames $(A, B, ...)$, then as above each one has a subsequence from χ. Each stratum i can be defined by taking the first n_i points in $\{y_j\}_{j=0}^{\infty}$ where n_i is the sample size required. Each strata then becomes its own BAS sample.

If unequal probability sampling is required another dimension can be added to BAS to achieve an arbitrary inclusion density function [17]. The same points in the 2-dimensional Master Sample will be used but with this extra dimension chosen by a new random integer. This changes how you sample from 2-dimensional space and thus the Master Sample points are no longer necessarily

chosen in the order that they appear in the original Master Sample.

2.5 Panel Designs

Panels are used to define a revisit strategy through time in order to balance status and trend monitoring. The combination of panels in each year needs to be representative as well as the entire sample through time. This is achieved by choosing the subset of points $\{y_j\}_{j=0}^{\infty}$ that fall within the sample frame and then selecting in order points for each panel matched to how they will be sampled in time. Thus, panel k with sample size n_k is $\{y_{a_k}, y_{a_k+1}, y_{a_k+2}, \ldots, y_{a+n_k-1}\}$ where $a_k = \sum_{i=1}^{k-1} n_i$. As these plots are set up each year points from the following panels can be included as replacement for sites that are actually not possible to sample. These points are then replaced the following panels until all panels are full. Removing points if budgets change should be done as last in, first out. Defining the panels this way the overall set is still a BAS design. Once a full rotation of all samples has been carried out, if additional points are needed a new panel is defined with the revisit schedule needed and these points are generated in order from the remaining points $\{y_j\}_{j \geq \sum_{\forall i} n_i}$. Table 1 shows an example panel design.

3 Application

3.1 Example 1: Tier 2 Monitoring Programme

3.2 Example 2: Abel Tasman Monitoring

4 Discussion

Benefits of BAS over GRTS as a master sample. Coordination of monitoring effort. Changing boundaries. Analysis.

References

[1] Roger S. Bivand, Edzer Pebesma, and Virgilio Gomez-Rubio. *Applied spatial data analysis with R, Second edition.* Springer, NY, 2013.

[2] D. A. Coomes, R. B. Allen, N. A. Scott, C. Goulding, and P. Beets. Designing systems to monitor carbon stocks in forests and shrublands. *Forest Ecology and Management*, 164(1):89–108, 2002.

[3] S. A. Field, P. J. O'Connor, A. J. Tyre, and H. P. Possingham. Making monitoring meaningful. *Austral Ecology*, 32(5):485–491, 2007.

[4] R. A. Gitzen. *Design and analysis of long-term ecological monitoring studies.* Cambridge University Press, 2012.

[5] A. Grafström, N. L. P. Lundström, and L. Schelin. Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2):514–520, 2012.

[6] T. M. Kincaid and A. R. Olsen. *spsurvey: Spatial Survey Design and Analysis*, 2016. R package version 3.3.

[7] A. J. King. The master sample of agriculture: I. development and use. *Journal of the American Statistical Association*, 40(229):38–45, 1945.

[8] D. P. Larsen, A. R. Olsen, and D. L., Jr. Stevens. Using a master sample to integrate stream monitoring programs. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(3):243–254, 2008.

[9] C. J. Legg and L. Nagy. Why most conservation monitoring is, but need not be, a waste of time. *Journal of environmental management*, 78(2):194–199, 2006.

[10] T. L. McDonald. Review of environmental monitoring methods: survey designs. *Environmental monitoring and assessment*, 85(3):277–292, 2003.

[11] J. D. Nichols and B. K. Williams. Monitoring for conservation. *Trends in ecology & evolution*, 21(12):668–673, 2006.

[12] Department of Conservation.

[13] Edzer J. Pebesma and Roger S. Bivand. Classes and methods for spatial data in R. *R News*, 5(2):9–13, November 2005.

[14] S. A. Peterson, N. S. Urquhart, and E. B. Welch. Sample representativeness: a must for reliable regional lake condition estimates. *Environmental science & technology*, 33(10):1559–1565, 1999.

[15] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[16] J. H Reynolds, K. B. Knutson, M. G .and Newman, E. D. Silverman, and W. L. Thompson. A road map for designing and implementing a biological monitoring program. *Environmental Monitoring and Assessment*, 188(7):1–25, 2016.

[17] B. L. Robertson, J. A. Brown, T. McDonald, and P. Jaksons. BAS: Balanced acceptance sampling of natural resources. *Biometrics*, 69(3):776–784, 2013.

[18] B. L. Robertson, T. McDonald, C. J. Price, and J. A. Brown. The halton frame: A useful sampling frame for balanced acceptance sampling. unpublished.

[19] J. R. Skalski. A design for long-term status and trends monitoring. *Journal of Environmental Management*, 30(2):139–144, 1990.
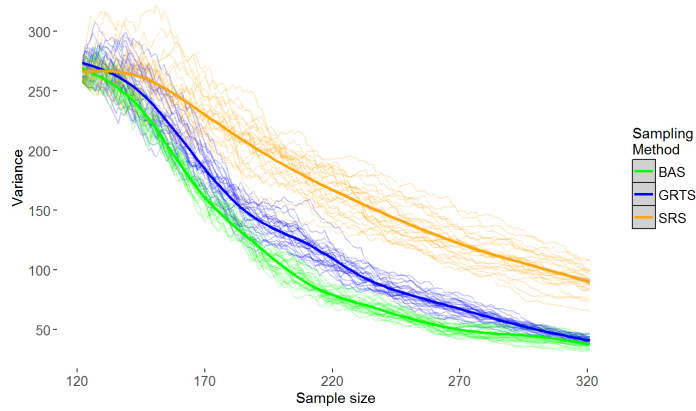
Figure 1: Results from a simulation study testing the spatial balance of 120 systematic points with BAS, GRTS, and Simple Random Sampling (SRS) points added on top.

[20] D. L., Jr. Stevens. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, 8(3):167–195, 1997.

[21] D. L., Jr. Stevens and A. R. Olsen. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99(465):262–278, 2004.

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

[23] P. Vos, E. Meelis, and W. J. Ter Keurs. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. *Environmental monitoring and assessment*, 61(3):317–344, 2000.

[24] N. G. Yoccoz, J. D. Nichols, and T. Boulinier. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, 16(8):446–453, 2001.
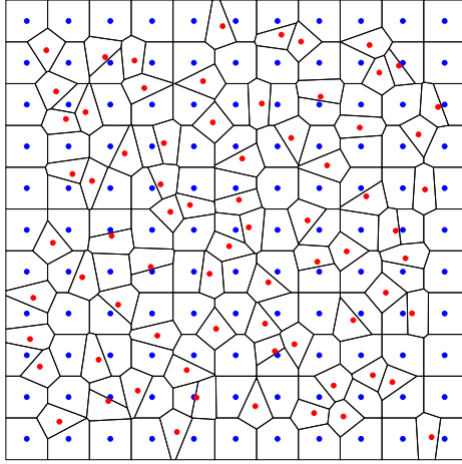
Appendix

Figure 2: Example of BAS Points being added to a systematic grid where blue is the existing points and red is BAS.

Panel	Sample Occasion									
	1	2	3	4	5	6	7	8	9	10
1	X	X	X	X	X	X	X	X	X	X
2	X			X			X			X
3		X			X			X		
4			X			X			X	

Table 1: Example of a panel design with revisit strategy of $[1-0, (1-2)^3]$.

Figure 3: A Halton grid with $(2, 3)$ co-prime base. The first six points in a sequence starting in B have the pattern (B,C,F,A,D,E).