

1 Web-based supplementary material for Using  
2 balanced acceptance sampling as a master sample  
3 for environmental surveys

4 Paul van Dam-Bates<sup>1,\*</sup>, Oliver Gansell<sup>2</sup>, and Blair Robertson<sup>3</sup>

5 <sup>1</sup>*Department of Conservation, Christchurch, New Zealand*

6 <sup>2</sup>*Department of Conservation, Hamilton, New Zealand*

7 <sup>3</sup>*University of Canterbury, Christchurch, New Zealand*

8 <sup>\*</sup>*Corresponding author: Paul van Dam-Bates, pbates@doc.govt.nz*

9 September 27, 2017

10 **1 Simulation Study**

11 We investigated estimation using the sampling methods balanced acceptance  
12 sampling (BAS), generalised random tessellation stratified (GRTS) and altered  
13 balanced acceptance sampling (aBAS) for augmenting legacy monitoring. To  
14 test differences between the methods, three response surfaces were generated: a

15 strong spatial trend (Population 1), a peak (Population 2), and a cyclical trend  
 16 (Population 3). The three functions used to define response values are listed  
 17 below.

- Population 1 (Robertson *et al.* 2013; Grafström *et al.* 2012):

$$f(\mathbf{x}) = 1000 [3(x_1 + x_2) + \sin(6(x_1 + x_2))],$$

18 with population total  $\tau \approx 2999.4$  (see Figure 1(a)).

- Population 2 (Peak function):

$$\begin{aligned} f(\mathbf{x}) = 10^6 [ & 3(4 - 6x_1)^2 \exp(-(6x_1 - 3)^2 - (6x_2 - 2)^2) \dots \\ & - 10(0.2(6x_1 - 3) - (6x_1 - 3)^3 - (6x_2 - 3)^5) \exp(-(6x_1 - 3)^2 - (6x_2 - 3)^2) \dots \\ & - \frac{1}{3} \exp(-(6x_1 - 2)^2 - (6x_2 - 3)^2) ], \end{aligned}$$

19 with population total  $\tau \approx 36270$  (see Figure 1(b)).

- Population 3 (Bird function):

$$\begin{aligned} f(\mathbf{x}) = 1000 [ & (12x_1 - 12x_2)^2 + \exp[(1 - \sin(12x_1 - 6))^2] \cos(12x_2 - 6) \dots \\ & + \exp[(1 - \cos(12x_2 - 6))^2] \sin(12x_1 - 6) ], \end{aligned}$$

20 with population total  $\tau \approx 23398.2$  (see Figure 1(c)).

21 The sampling frame was defined as  $100 \times 100$  raster in  $[0, 1)^2$ . The response  
 22 value for each raster cell was defined as the integral of  $f(\mathbf{x})$  over the cell. Sce-  
 23 narios similar to Foster *et al.* 2017 using the program R (R Core Team 2015)  
 24 were run. We assumed an arbitrary overall sample size of  $n = 60$ . Legacy  
 25 plots ( $n_l \in 3, 4, \dots, 57$ ) were generated either as simple random samples (SRS)

26 or random-start systematic sampling (SS). More samples ( $n_b = 60 - n_l$ ) were  
 27 then included using GRTS (Kincaid & Olsen 2016), BAS, and aBAS (Foster  
 28 2016). Simple random samples were added to SS legacy plots as well.

Each raster cell has inclusion probability of  $\pi = \frac{n}{100^2}$  when using equal probability sampling. Analysis was carried out for BAS and GRTS as

$$\bar{y} = \frac{1}{10000} * \sum_{i=1}^{60} y_i / \pi_i.$$

Where  $y_i$  is the observed response value for sample  $i$ . For aBAS we followed as described in Foster et al. (2017) and calculated the sample mean as

$$\bar{y} = \frac{1}{10000} * \left( \frac{n_l}{60} \sum_{i=1}^{n_l} y_i / \pi_l + \frac{n + n_l}{60} \sum_{i=1}^n y_i / \pi_{ai} \right).$$

29 For each sample size of  $3 \leq n_l \leq 57$  we ran 1000 simulations.

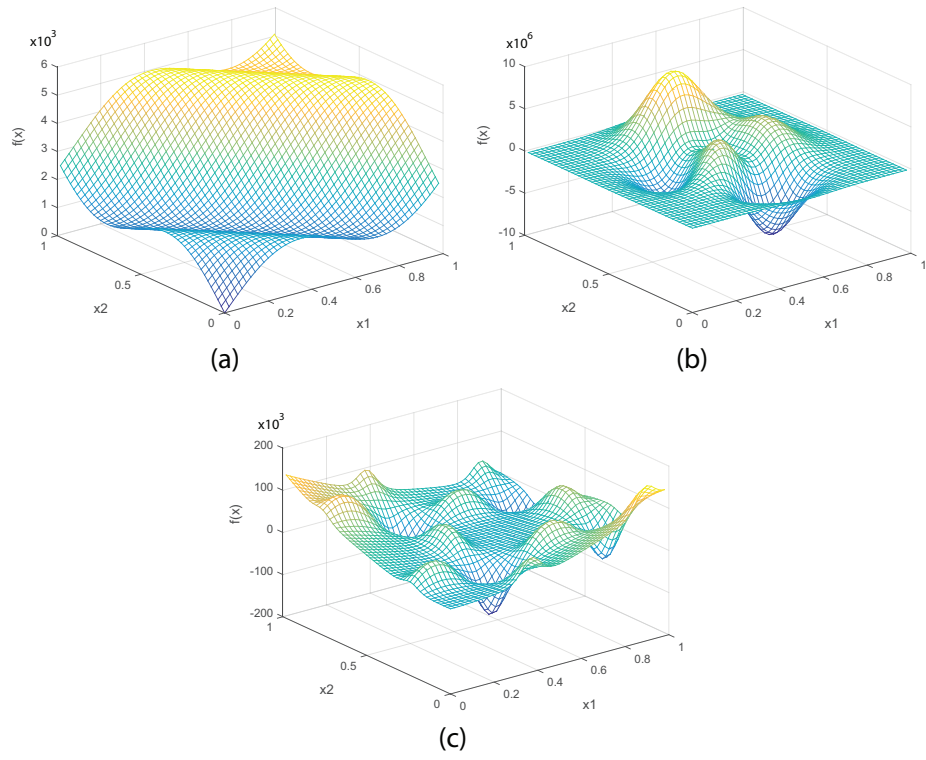


Figure 1: The three populations used to test estimation in Section 2.6. (a) Population 1, (b) Population 2 and (c) Population 3.

## 30 2 Master Sample R Code (NZ)

```

31
32 library(sp)
33
34 #Halton Sequence:
35 RSHalton <- function(n = 10,seeds = c(0,0),bases = c(2,3)) {
36     ##
37     ## Generate n points from a random start d dimensional
38     ↪ Halton sequence.
39     ##
40     ## Inputs:
41     ##
42     ## n      sample size
43     ## bases  coprime bases e.g. c(2,3) (Halton Sequence)
44     ## seeds  random seeds e.g. c(0,0) (Halton Sequence)
45
46
47     ##### Initialize
48     ↪ #####
49     d <- length(bases); pts <- mat.or.vec(n, d)
50     if (length(seeds) != d){
51         seeds <- rep(seeds[1],d)
52     }

```

```

53
54     ##### Main Loop
55     ↪ #####
56     for (i in 1:d) {
57         b <- bases[i]; u <- seeds[i]; k <- u:(u+n-1);
58         xk <- (k %% b)/b;
59         for (j in 1:(ceiling(logb(u+n,b)) + 2)) {
60             xk <- xk + (floor(k/(b^j)) %% b)/(b^(j+1));
61         }
62         pts[,i] <- cbind(xk)
63     }
64     pts <- cbind(1:nrow(pts), pts)
65     return(pts)
66 }
67
68 #Generate the master sample after the seeds have been manually
69     ↪ set.
70 masterSample <- function(island = "South", shp, N = 100){
71     #Master Sample seed for South Island, chosen as first random
72     ↪ start that fell into SI
73     #seed.si <- c(4887260, 18041662)
74     #seed.ni <- c(5137598, 8906854)
75

```

```

76 #Define CRS
77 nztm <- "+proj=tmerc_+lat_0=0_+lon_0=173_+k=0.9996_+x_0=1600000_+y
78     ↪ _0=10000000_+ellps=GRS80_+towgs84=0,0,0,0,0,0,0_+units=m_+
79     ↪ no_defs"
80
81 if(island == "South")
82 {
83   bb <- data.frame(min = c(1089354,4747979), max = c
84     ↪ (1721164,5516919), row.names = c("x","y"))
85   seed <- c(4887260, 18041662)
86 }else if(island == "North")
87 {
88   bb <- data.frame(min = c(1510593,5390569), max = c
89     ↪ (2092000,6223164), row.names = c("x","y"))
90   seed <- c(5137598, 8906854)
91 }else{
92   cat("ERROR: Define Island for MS\n")
93   return()
94 }
95
96 #Scale and shift Halton to fit into bounding box
97 scale.bas <- bb[,2] - bb[,1]
98 shift.bas <- bb[,1]

```

```

99
100 draw <- 10000
101
102 getSample <- function(k = 0){
103   if(k == 0){ seedshift <- seed
104   }else seedshift <- k*draw + seed
105   pts <- RSHalton(n = draw, seeds = seedshift, bases = c(2,3))
106   pts[,2] <- pts[,2]*scale.bas[1] + shift.bas[1]
107   pts[,3] <- pts[,3]*scale.bas[2] + shift.bas[2]
108
109   #Give points a projection, clip them as needed.
110   tmp.order <- (k*draw + 1):((k+1)*draw)
111   pts.coord <- SpatialPointsDataFrame(cbind(pts[,2],pts[,3]),
112     ↪ proj4string=CRS("+proj=tmerc_+lat_0=0_+lon_0=173_+k=0.9996
113     ↪ _+x_0=1600000_+y_0=10000000_+ellps=GRS80_+units=m_+no_defs
114     ↪ "), data.frame(SiteOrder = tmp.order))
115   return(pts.coord)
116 }
117
118 pts.sample <- getSample()
119 pts.sample <- pts.sample[shp, ]
120
121

```



```

122 if(nrow(pts.sample) < N){
123   di <- 1
124   while(nrow(pts.sample) < N){
125     new.pts <- getSample(k = di)
126     new.pts <- new.pts[shp, ]
127     pts.sample <- rbind(pts.sample, new.pts)
128     di <- di + 1
129   }
130   return(pts.sample[1:N,])
131 } else{
132   return(pts.sample[1:N,])
133 }
134 }
135

```

## 136 References

- 137 Foster, S.D. (2016) *MBHdesign: Spatial Designs for Ecological and Environ-*  
138 *mental Surveys*. R package version 1.0.61.
- 139 Foster, S.D., Hosack, G.R., Lawrence, E., Przeslawski, R., Hedge, P., Caley,  
140 M.J., Barrett, N.S., Williams, A., Li, J., Lynch, T. *et al.* (2017) Spatially-  
141 balanced designs that incorporate legacy sites. *Methods in Ecology and Evo-*  
142 *lution*.

- 143 Grafström, A., Lundström, N.L.P. & Schelin, L. (2012) Spatially balanced sam-  
144 pling through the pivotal method. *Biometrics*, **68**, 514–520.
- 145 Kincaid, T.M. & Olsen, A.R. (2016) *spsurvey: Spatial Survey Design and Anal-*  
146 *ysis*. R package version 3.3.
- 147 R Core Team (2015) *R: A Language and Environment for Statistical Computing*.  
148 R Foundation for Statistical Computing, Vienna, Austria.
- 149 Robertson, B.L., Brown, J.A., McDonald, T. & Jaksons, P. (2013) BAS: Bal-  
150 anced acceptance sampling of natural resources. *Biometrics*, **69**, 776–784.