

## Sample Design for Kahna Pench, India Bioacoustic Study

For identifying a field-based sample distribution for testing variation across a disturbance gradient, using bioacoustic sensors, we implemented a random weighted stratified design with a 600m minimum sampling distance. Eight unique strata were defined by disturbance quantiles, forest/non-forest and energy development “go” areas. Each stratum had 4 replicates located with 4 block sample replicates, for use as additional plots if plots in the previous block are found to be unsuitable. Sample weights were used in the randomization, based on an inverted  $[(((x - \max(x)) * -1) + \min(x))]$  and transformed  $[x / \sum(x)]$  disturbance index, to optimize across the disturbance gradient. The nested sampling approach (Schlup & Wagner 2008; Funk et al., 2016) is implemented to: 1) optimize statistical power while maintaining observational independence 2) and capture distance (spatial-lag) and directional (anisotropy) spatial influences. Two distance bandwidths (200m-400m and 400m-600m) were selected, to generate plots at random distance and bearing, based on experimental testing of distance/die-off of the acoustic sensors recording +/-200m (personal comm Columbia University). Whereas these plots are “clustered” statistically, they would be considered independent experimental units thus, not unduly effecting the degrees of freedom issues that often plague transect based designs.

Starting minimum sample sizes per stratum we calculated using a power test (Lipsey 1990). This represented the pool of potential sample locations that were randomized for filling the replicates in each stratum. The optimal number of replicates, in each stratum, were modeled using a jackknife of effect size (Cohen 1988). We found that just the stratified plot locations were not adequate to characterize the disturbance gradient but (table 2) however, when combined with the nested plot approach the effect sizes went up significantly (all in the very-large to huge range). We also tested a  $n = [1-6]$  range of replicated stratum plots and found that 4 was the optimal number of plots to positively influence all pair-wise effect sizes (table 1) without unduly inflating the overall sample size.

The sampling strategy follows:

- 1) Calculate approximate number of samples using  $[n = (2 * \delta(y))^2 / \delta_x(y)]$  based on the disturbance index (Lipsey 1990), per strata.
- 2) Create stratified random sample using transformed inverted disturbance (as probabilistic sampling weights) conditional on strata defined by non-forest, go areas and quantiles of disturbance, for 8 unique strata combinations ( $s=8$ ), with a 600m minimum distance constraint.
- 3) Once the stratified samples are generated, apply the nested sampling approach where two additional random samples are created at random bearing/distance, at two distance lags (200m-400m and 400m-600m), for each stratified sample.
- 4) Finally, accounting for number of replicates per strata, iteratively apply a step-down jackknife power analysis on the resulting sample pool using Cohen’s effect size (Cohen 1988) as the evaluation statistic until the statistic ( $d$ ) is rejected at  $[d < 0.5]$ . This will be our final supported sample size. We will test this with and without the inclusion of the cluster, distance lagged, plots.

Effect size ( $d$ ) is defined as;

$$d = \frac{\mu_1 - \mu_2}{s};$$
$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Where;  $s$  = pooled standard deviation,

$s_1^2$  and  $s_2^2$  are the standard deviations for the groups.

## References

- Cohen, J., (1988) Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates.
- Lipsey, Mark (1990). *Design Sensitivity: Statistical Power for Experimental Research*; Sage Publications.
- Schlup, B.M., H.H Wagner (2008) Effects of study design and analysis on the spatial community structure detected by multiscale ordination. *Journal of Vegetation Science*, 19(5):621-632
- Funk, W.C., M.A. Murphy, K.L. Hoke, E. Muths, S.M. Amburgey, E.M. Lemmon, A.R. Lemmon, J.S. Evans (2016) Elevational speciation in action? Restricted gene flow associated with adaptive divergence across an altitudinal gradient *Journal of Evolutionary Biology*, 29(2):241-252

strata pairs	lower	effect	upper
8 and 9	0.204	1.154	2.066
8 and 10	1.215	2.560	3.864
8 and 11	1.563	3.083	4.566
8 and 12	-4.604	-3.112	-1.581
8 and 13	-4.418	-2.974	-1.491
8 and 15	-10.070	-7.064	-4.040
8 and 16	-10.199	-7.156	-4.095
9 and 10	0.715	1.839	2.920
9 and 11	1.237	2.593	3.907
9 and 12	-8.218	-5.737	-3.232
9 and 13	-6.509	-4.504	-2.469
9 and 15	-13.671	-9.633	-5.584
9 and 16	-13.333	-9.392	-5.440
10 and 11	0.748	1.886	2.980
10 and 12	-12.612	-8.879	-5.132
10 and 13	-8.605	-6.014	-3.402
10 and 15	-16.694	-11.920	-6.948
10 and 16	-15.852	-11.314	-6.587
11 and 12	-7.751	-5.400	-3.025
11 and 13	-7.449	-5.183	-2.891
11 and 15	-11.602	-8.158	-4.700
11 and 16	-11.778	-8.284	-4.775
12 and 13	-1.535	-0.698	0.166
12 and 15	-7.894	-5.503	-3.088
12 and 16	-8.094	-5.647	-3.177
13 and 15	-5.251	-3.588	-1.890
13 and 16	-5.701	-3.917	-2.099
15 and 16	-1.414	-0.589	0.259

Table 1. Pairwise strata Cohen's  $d$  effect size(s) for all plots (clusters)  $n=96$ .

strata pairs	lower	effect	upper
8 and 9	-0.821	-0.021	0.780
8 and 10	-1.350	-0.531	0.310
8 and 11	-1.128	-0.324	0.494
8 and 12	-2.179	-1.247	-0.276
8 and 13	-1.930	-1.040	-0.114
8 and 15	-1.850	-0.972	-0.059
8 and 16	-2.038	-1.131	-0.185
9 and 10	-1.225	-0.416	0.411
9 and 11	-1.035	-0.234	0.577
9 and 12	-1.703	-0.846	0.043
9 and 13	-1.579	-0.738	0.133
9 and 15	-1.588	-0.746	0.126
9 and 16	-1.658	-0.807	0.075
10 and 11	-0.608	0.201	1.001
10 and 12	-1.103	-0.300	0.516
10 and 13	-1.026	-0.226	0.585
10 and 15	-1.082	-0.280	0.535
10 and 16	-1.094	-0.291	0.524
11 and 12	-1.405	-0.582	0.266
11 and 13	-1.300	-0.485	0.350
11 and 15	-1.335	-0.517	0.322
11 and 16	-1.377	-0.556	0.288
12 and 13	-0.722	0.081	0.880
12 and 15	-0.818	-0.018	0.783
12 and 16	-0.804	-0.004	0.796
13 and 15	-0.884	-0.085	0.718
13 and 16	-0.879	-0.080	0.722
15 and 16	-0.787	0.013	0.813

Table 2. Pairwise strata Cohen's  $d$  effect size(s) for center plots (main stratifying plot)  $n=32$ .

## R code for stratification

---

```
#####  
#### Sample design for Kahna Pench, India bioacoustic study  
#####  
library(sp)  
library(raster)  
library(spatialEco)  
library(rgdal)  
  
# Working (path) and data directories (dpath)  
path = "C:/evans/India/kahna_pench/data"  
dpath = "C:/evans/India/kahna_pench/raw_data/"  
setwd(path)  
  
utm.proj <- "+proj=utm +zone=43 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0"  
out.shape = "stratification_v2" # name of resulting shapefile  
p = 600 # minimum sample distance  
min.dist = 200 # minimum distance for lag sample  
max.dist = 600 # maximum distance for lag sample  
sn = 4 # Number of random samples per strata  
sr = 4 # Number of strata replicates  
  
# function for standard error  
std.err <- function(x, na.rm = TRUE) {  
  if(na.rm) {  
    x <- x[!is.na(x)]  
    return( sd(x) / sqrt(length(x)) )  
  } else {  
    return( NA )  
  }  
}  
  
#####  
# Read, crop and mask stratification data  
#####  
# Stratification (strat.tif)  
# * strata used in stratification  
# strat disturbance forest go/no  
# 1 1 1 0  
# *2 1 0 0  
# 3 2 1 0  
# 4 3 1 0  
# *5 3 0 0  
# 6 4 1 0  
# *7 4 0 0  
# *8 3 0 1  
# 9 3 1 1  
# 10 4 1 1  
# *11 4 0 1  
# *12 2 0 1  
# 13 2 1 1
```

```

# *14      2      0      0
# *15      1      0      1
# 16      1      1      1

# Read stratification raster and table
r <- raster("strat.tif")
dat <- read.csv("strat_table.csv")
dat <- dat[,-2]

# Index "forested" strata and remove
# strata raster
fidx <- dat[, "value"][which(dat$forest == 1)]
r[which(r[] %in% fidx)] <- NA

# Number of unique strata
cat("Number of unique strata", length(unique(r[])) - 1, "\n")

# Read forest/non-forest raster
f <- raster(paste0(dpath, "forest_ANYeq1.tif"))
f <- mask(f, crop(f, extent(r), snap = "in"), r)

# Read disturbance raster
d <- raster(paste0(dpath, "di3_sum3_pta_mahmp_1ki.tif"))
d <- mask(crop(d, extent(r)), r)
#d <- d / 1000

#####
# sample non-forest in go and no-go
# using disturbance as sample weights
#####

strata.ids <- as.vector(na.omit(unique(r[])))

# Power-based minimal sample size for each strata
n = vector()
for( i in strata.ids ) {
  n <- append(n, round((( 2 * sd(d[which(r[] == i)],
    na.rm = TRUE))^2 /
    std.err(d[which(r[] == i)])) , 0))
}
names(n) <- strata.ids
n <- round(n / 20, 0)
# Draw stratified random samples
strata <- data.frame(matrix(vector(), 0, 3,
  dimnames=list(c(),
    c("x", "y", "strata"))))
for( i in strata.ids ) {
  idx <- sample(which(r[] == i),
    size = (n[which(names(n) == i)]),
    replace = FALSE)
  cat("Pulling", length(idx), "random samples for strata", i, "\n")
  strata <- rbind(strata, data.frame(xyFromCell(r, idx),
    strata = i) )
}

```

```

}
# Coerce results to sp SpatialPointsDataFrame object
coordinates(strata) <- ~x+y

# Count number of obs in each strata
for( i in strata.ids ) {
  cat("Obs in strata: ", i,
      length(strata[strata$strata == i,$strata]), "\n")
}

# N per strata 4; reps per N, per strata 4
strata.n4r4 <- stratified.random(strata, strata = "strata",
                                n = sn, reps = sr)
strata.sub <- strata.n4r4[strata.n4r4$REP == 1,]
rad.list <- list()
for(i in 1:nrow(strata.sub)) {
  d1 <- sample.annulus(strata.sub[i,], r1=min.dist, r2=(min.dist+min.dist),
                      n = 1, type = "random", iter=50)
  d1$rmin <- min.dist
  d1$rmax <- (min.dist + min.dist)
  d2 <- sample.annulus(strata.sub[i,], r1=(min.dist + min.dist), r2=max.dist,
                      n = 1, type = "random", iter=100)
  d2$rmin <- (min.dist + min.dist)
  d2$rmax <- max.dist
  rad.list[[i]] <- rbind(d1,d2)
}

# Format final stratification SpatialPointsDataFrame
cluster.plots <- do.call("rbind", rad.list)
strata.sub@data <- data.frame(SID=as.numeric(row.names(strata.sub@data)),
                             rmin=0, rmax=0,s=strata.sub$strata)
cluster.plots <- merge(cluster.plots, strata.sub@data[,c("SID", "s")], by="SID")
cluster.plots@data <- cluster.plots@data[,-2]

strata.final <- rbind(strata.sub,cluster.plots)
strata.final$strata <- as.numeric(as.character(strata.final$s))
strata.final@data <- strata.final@data[,-which(names(strata.final) == "s")]
strata.final <- strata.final[order(strata.final$strata, strata.final$SID, strata.final$rmin,
                                decreasing = c(FALSE,FALSE,FALSE)),]

# Write stratification shapefile
proj4string(strata.final) <- utm.proj
writeOGR(strata.final, getwd(), out.shape, driver="ESRI Shapefile",
        check_exists=TRUE, overwrite_layer=TRUE)

```

```
#####
# Test effect size
#####
d <- raster(paste0(dpath,"di3_sum3_pta_mahmp_1ki.tif"))
strata.final$disturb <- extract(d, strata.final)/1000

effect.size <- list()
cbd <- combn(unique(strata.final$strata), 2)
for(i in 1:ncol(cbd)) {
  es.test <- data.frame(strata = strata.final[strata.final$strata == cbd[i][1] |
    strata.final$strata == cbd[i][2]],)$strata,
    y = strata.final[strata.final$strata == cbd[i][1] |
    strata.final$strata == cbd[i][2]]$disturb)
  effect.size[[ paste0(cbd[i][1], "_", cbd[i][2]) ]] <- psych::cohen.d(es.test, "strata")$cohen.d
}
(effect.size <- data.frame(strata = names(effect.size),
  do.call("rbind", effect.size)))
write.csv(effect.size, "effect_size.csv", row.names=FALSE)

effect.size.plot <- list()
strata.plot <- strata.final[strata.final$rmin ==0,]
cbd <- combn(unique(strata.plot$strata), 2)
for(i in 1:ncol(cbd)) {
  es.test <- data.frame(strata = strata.plot[strata.plot$strata == cbd[i][1] |
    strata.plot$strata == cbd[i][2]],)$strata,
    y = strata.plot[strata.plot$strata == cbd[i][1] |
    strata.plot$strata == cbd[i][2]]$disturb)
  effect.size.plot[[ paste0(cbd[i][1], "_", cbd[i][2]) ]] <- psych::cohen.d(es.test, "strata")$cohen.d
}
(effect.size.plot <- data.frame(strata = names(effect.size.plot),
  do.call("rbind", effect.size.plot)))
write.csv(effect.size.plot, "effect_size_plot.csv", row.names=FALSE)
```