



POS 1 – PREPARATION DE LA CONCEPTION DE L'ECHANTILLONNAGE

Laboratoire d'Observation des Forêts de Madagascar (LOFM), 2021

TABLE DES MATIERES

1.	Information de base.....	1
1.1.	Objectifs.....	1
1.2.	Responsabilités.....	1
1.3.	Conditions préalables.....	1
1.4.	Documents connexes.....	1
2.	Procédure.....	2
2.1.	Étape 0 : Sondage pilote.....	2
2.2.	Étape 1 : Déterminer la conception de base de l'échantillonnage.....	2
2.3.	Étape 2 : Déterminer la stratification.....	3
2.4.	Étape 3 : Établissement du nombre d'unités d'échantillonnage.....	5
2.4.1.	Calcul des superficies par strate.....	7
2.4.2.	Calcul de la taille de l'échantillonnage.....	13
2.5.	Étape 4 : Sélection des unités d'échantillonnage.....	13
2.6.	Important: Revue de l'échantillonnage et finalisation de la stratification.....	20

LISTE DES CARTES

Carte 1 : Exemple de carte de stratification pour l'échantillonnage (avec buffer) (à gauche) Vs Carte initiale présentant les changements d'utilisations des terres ou les strates respectant les UOT.....	4
--	---

LISTE DES EQUATIONS

Équation 1: Erreurs types attendues et incertitudes en pourcentage attendues pour le nombre d'unités d'échantillonnage alloué dans chaque strate.....	5
Équation 2: Taille totale de l'échantillon.....	6

1. INFORMATION DE BASE

1.1. Objectifs

Cette Procédure d'Opération Standard (POS) sert à établir une conception d'échantillonnage spatialement référencée, fondée sur les probabilités et géographiquement équilibrée pour l'estimation des zones dans les relevés terrestres. Elle est applicable pour le suivi avec un échantillonnage stratifié.

Remarque : Le cas de suivi à l'aide de grille systématique, à l'exemple de la grille nationale de référence de 4 km x 4 km de Madagascar : outil pour le suivi et l'inventaire forestier est expliqué dans un autre Document.

1.2. Responsabilités

Les responsables de Laboratoire et de la Méthodologie seront chargés de **discuter de la bonne conception de l'échantillonnage**.

Les analystes/techniciens de Laboratoire seront responsables de l'affinement de la conception d'échantillonnage en discutant avec les Responsables ; et **chargés des aspects statistiques de l'enquête : détermination du nombre d'unités échantillonnées, calcul des erreurs... ».**

1.3. Conditions préalables

Si l'**échantillonnage stratifié** est sélectionné, **des cartes devront être générées pour chaque période de surveillance en tant que cadre de stratification**.

1.4. Documents connexes

- *POS0 : Cartographie de la stratification*
- *POS2 : Conception de la réponse*
- *POS3 : Collecte de données*
- *POS4 : Analyse des données*
- Laboratoire d'Observation des Forêts de Madagascar (LOFM), Bureau National des Changements Climatiques et de la Réduction des Émissions dues à la Déforestation et à la Dégradation des Forêts (BNCCREDD+) - Ministère de l'Environnement et du Développement Durable. 2019. Directives d'utilisation de la grille nationale de référence (4 km x 4 km), outil pour la collecte de données nationales d'inventaire forestier. 11. Antananarivo. Madagascar
- Bureau National de Coordination de la REDD+, Ministère de l'Environnement, de l'Ecologie, de la Mer et des Forêts. 2018. Assurance de la qualité. 9. Antananarivo. Madagascar
- Bureau National de Coordination de la REDD+, Ministère de l'Environnement, de l'Ecologie, de la Mer et des Forêts. 2018. Collecte de données. 26. Antananarivo. Madagascar
- Bureau National de Coordination de la REDD+, Ministère de l'Environnement, de l'Ecologie, de la Mer et des Forêts. 2018. Protocole de classification. 26. Antananarivo. Madagascar

2. PROCÉDURE

2.1. Étape 0 : Sondage pilote

La première étape de l'échantillonnage est de générer une **étude pilote** pour connaître les systèmes ou plans d'échantillonnage adaptés à l'étude. Des études pilotes peuvent ainsi être faites en première étape de ce SOP0.

Les responsables de Laboratoire et de la Méthodologie, en coordination avec les analystes/techniciens du Laboratoire, déterminent la nécessité de mener une enquête pilote qui servira à éclairer le plan d'échantillonnage.

Les responsables du Laboratoire et de la Méthodologie avec les analystes suivent les étapes 1 à 4 et les SOP 2-4 pour mettre en œuvre l'enquête pilote et documentent les leçons tirées des résultats à l'aide du Modèle 1 (en Annexe).

Pour l'échantillonnage stratifié, **environ 100 unités d'échantillonnage par strate** sont sélectionnées. Les unités d'échantillonnage sont ensuite évaluées en suivant les POS2, POS3 et POS4. **La proportion des variables d'intérêt par strate sera obtenue**, et il sera possible de déterminer la taille de l'échantillon avec ces informations.

Remarque : Une autre option consiste à évaluer les points de 4km x 4km de la grille nationale de référence et utiliser cela comme pilote, cette option présente l'avantage de pouvoir faire un suivi annuel sur les points permanents de la grille nationale.

2.2. Étape 1 : Déterminer la conception de base de l'échantillonnage

Cette étape consiste à déterminer le type d'échantillonnage adapté sur la base des variables étudiées et des principes de statistique existantes.

Pour le suivi du programme de réduction d'émissions (et même les initiatives), l'échantillonnage est stratifié car la période de suivi est courte (1-3 ans). L'échantillonnage systématique est utilisé pour le niveau de référence et aussi les périodes plus longues (5-10 ans).

Les responsables du Laboratoire et de la Méthodologie, avec les analystes/techniciens du laboratoire déterminent **le plan d'échantillonnage de base**, y compris **le type d'échantillonnage** (STR stratifié, SYS systématique ou SRS aléatoire simple), **le type d'unité d'échantillonnage** (pixel de la carte, points) et **la forme et taille du support spatial utilisé pour l'interprétation, y compris la définition de toute sous-unité d'échantillonnage**.

Les responsables du Laboratoire et de la Méthodologie avec les analystes/techniciens de Laboratoire doivent documenter la justification du plan d'échantillonnage choisi à l'aide du Modèle 1.

Remarque : Si le plan d'échantillonnage sélectionné est différent des plans d'échantillonnage précédents utilisés pour la période de référence ou la période de surveillance précédente, la documentation doit également traiter de la manière dont le changement de conception affecte la comparabilité et l'exactitude des résultats. La justification doit être documentée et stockée dans le dossier MNV du Programme (base de données).

Pour notre cas c'est-à-dire le suivi du programme de réduction d'émissions et des initiatives, l'échantillonnage stratifié sera choisi. En effet, ce type d'échantillonnage est adapté pour une période de suivi qui est courte (1-3 ans).

L'échantillonnage systématique est utilisé pour le niveau de référence et aussi pour des périodes plus longues (5-10 ans).

2.3. Étape 2 : Déterminer la stratification

Les responsables du Laboratoire et de la Méthodologie en coordination avec les analystes/techniciens du laboratoire déterminent la stratification à utiliser en fonction des classes d'intérêt dans la zone.

Les responsables du Laboratoire et de la Méthodologie documentent la stratification à l'aide du Modèle 1.

Au minimum, cela inclut :

- **(i) le nombre de strates (H) ;**
- **(ii) une description de chaque strate indiquant le nom de la strate et une description ou une définition.**

Les analystes/techniciens du Laboratoire en coordination avec les responsables du Laboratoire et de la Méthodologie génèrent **la carte de stratification** qui divise la région d'intérêt en strates discrètes et non chevauchantes tout en assurant une représentation géographiquement équilibrée. Le responsable de Laboratoire fournit les instructions pour générer la carte de stratification. Les instructions peuvent indiquer l'utilisation de produits cartographiques spécifiques à utiliser, y compris la manière dont cela doit être appliqué.

Remarque : En cas de post-stratification, le responsable du Laboratoire avec les analystes/techniciens doivent documenter et stocker les critères de post-stratification sous le Modèle 1 stocké dans le dossier MNV du Programme (base de données).

Il est à préciser alors dans notre cas que :

- ⇒ La stratification décrite dans cette étape 2 est applicable uniquement dans le cas d'un **échantillonnage stratifié.**
- ⇒ Les strates sont définies à l'issue de la création de la carte de stratification (voir POS0).
- ⇒ Les strates utilisées sont :
 - Strate de forêt stable ou forêts persistantes : FF (forêt restant forêt), codé « 11 » ;

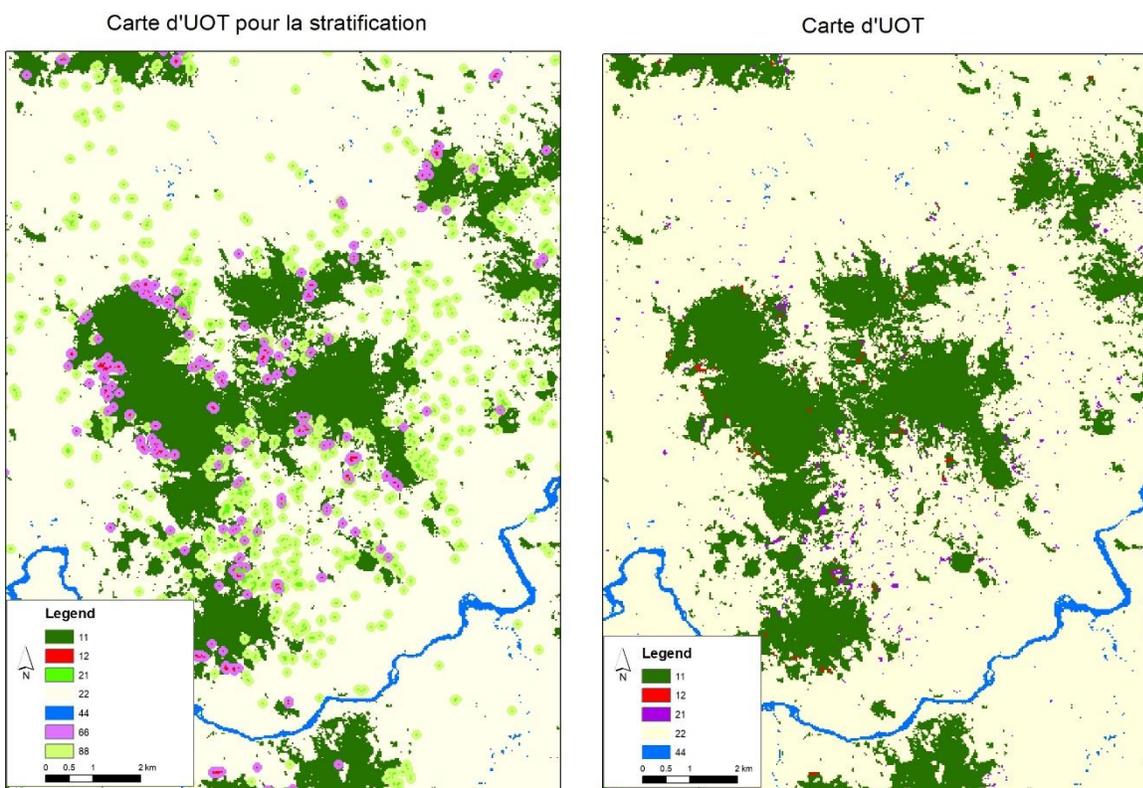
- Strate de déforestation : FN (forêt convertie en non forêt) sur laquelle la précision la plus élevée sera toujours recherchée durant la création de la carte de stratification, et qui est codé « 12 » ;
- Strate de non forêt persistantes : NN (non forêt restant non forêt), codé « 22 » ;

En plus de ces trois (03) strates, deux (02) classes sont ajoutées pour améliorer l'échantillonnage :

- Une première zone tampon de 50 m pour la forêts stable (vers l'intérieur) codée « 55 » et
- Une deuxième zone tampon de 50 m pour la forêt stable (vers l'intérieur, à partir de la strate « 55 ») codée « 56 ».

L'idée est que les erreurs de non inclusion de la déforestation peuvent être importantes autour de la classe de forêt stable et on voudrait donc les échantillonner proprement.

Avant l'échantillonnage proprement dit, on devra donc avoir **cinq (05) strates**.



Carte 1 : Exemple de carte de stratification pour l'échantillonnage (avec buffer) (à gauche) Vs Carte initiale présentant les changements d'utilisations des terres ou les strates respectant les UOT

La carte initiale devra ressembler à la **Erreur ! Source du renvoi introuvable.** (vue à droite), comparée avec la carte finale (vue à gauche), qui est celle qu'on crée et qui contient **les 05 strates**.

Les surfaces des 03 premières strates seront les surfaces réelles dans la carte UOT (à gauche).

Les surfaces des deux zones tampons seront calculées dans la carte finale.

L'eau ne sera pas échantillonnée, et donc quelle que soit la surface, le nombre d'échantillons sera égal à 0.

Les étapes pour la création de la carte de stratification sont détaillées dans le POS0.

2.4. Étape 3 : Établissement du nombre d'unités d'échantillonnage

Il s'agit d'une **sous étape applicable (comme pour l'étape 2) uniquement pour un échantillonnage stratifié** (par opposition à l'échantillonnage systématique).

Les analystes/techniciens du Laboratoire en coordination avec les responsables du Laboratoire et de la Méthodologie calculent **la taille totale de l'échantillon** en utilisant un processus itératif qui permet d'atteindre **une taille d'échantillon globale réalisable compte tenu du temps et des ressources disponibles et pour chaque variable d'intérêt ou variable clé d'intérêt.**

Remarque : Lorsque l'échantillonnage stratifié n'est pas sélectionné à l'étape 1, le nombre de strates est $H = 1$ et les équations se simplifient en conséquence.

Les analystes/techniciens du Laboratoire en coordination avec les Responsables Laboratoire et Méthodologie estiment **la superficie attendue de chaque strate à partir de la carte de stratification, les proportions de zone attendues sur la base d'un jugement sur l'occurrence probable des variables d'intérêt dans la zone ou sur la base de l'enquête pilote menée à l'étape 0.**

Les analystes/techniciens du laboratoire en coordination avec les responsables du Laboratoire et de la Méthodologie estiment **les erreurs types attendues et les incertitudes en pourcentage attendues pour le nombre d'unités d'échantillonnage alloué dans chaque strate à l'aide des équations suivantes :**

Équation 1: Erreurs types attendues et incertitudes en pourcentage attendues pour le nombre d'unités d'échantillonnage alloué dans chaque strate

$$S(p_h) = \frac{\sqrt{p_h * (1 - p_h)}}{n_h - 1}$$

$$U\%(p_h) = t_{a,df} * \frac{S(p_h)}{p_h}$$

$$U\%(p) = \frac{t_{a,df}}{p} * \sum_{h=1}^H W_h^2 * S^2(p_h)$$

Où

p est la proportion de surface à estimer pour la région d'intérêt
 ph est la proportion de surface à estimer pour la strate h
 S (ph) est l'erreur standard de la proportion de surface pour la strate h
 nh est le nombre d'unités d'échantillonnage dans la strate h
 Wh est le poids de la strate h - à estimer comme la proportion de la superficie totale dans la strate h
 U% (ph) est le pourcentage d'incertitude de la proportion de surface pour la strate h
 U% (p) est le pourcentage d'incertitude de la proportion de surface à estimer
 $t_{\alpha, df}$ est le niveau de confiance de Student t à un niveau $1-\alpha$ et le degré de liberté, $df = n_h - H - 1$.

En considérant que n soit grand, $df \rightarrow \infty$, et à 95% de niveau de confiance ; $t_{.05, \infty} \approx 1.96$ et à 90% de niveau de confiance $t_{.1, \infty} \approx 1.645$.

Le niveau de confiance à utiliser est de 0,95, soit t égal à 1.96.

Des scripts R (logiciel R) et Microsoft Office Excel sont utilisés pour produire les calculs.

Les analystes/techniciens du Laboratoire **résumant les proportions attendues, les erreurs standard et les incertitudes en pourcentage associées** avec le Modèle 1.

Les analystes/techniciens du laboratoire en coordination avec les responsables du Laboratoire et de la Méthodologie calculent **la taille totale de l'échantillon** à l'aide de l'équation ci-dessous, et modifient de manière itérative les variables attendues, principalement la marge d'erreur admissible d, pour atteindre une taille d'échantillon globale réalisable compte tenu du temps et des ressources disponibles.

Les responsables du laboratoire et de la méthodologie résumant cela dans le modèle 1.

Équation 2: Taille totale de l'échantillon

$$n \approx \left(\frac{t_{\alpha, df} \cdot \sum_{h=1}^H W_h \cdot S_h}{d} \right)^2$$

Le logiciel et scripts R et les feuillets Excel sont utilisés pour faire le calcul.

Les analystes/techniciens du Laboratoire en coordination avec les responsables du Laboratoire et de la Méthodologie définissent **les règles d'attribution des unités d'échantillonnage** à chaque strate (précisé et résumé dans le Modèle 1 si besoin).

Les analystes/techniciens du laboratoire en coordination avec les responsables du Laboratoire et de la méthodologie **ajustent le nombre d'unités d'échantillonnage dans chaque strate jusqu'à atteindre un pourcentage d'incertitude attendu satisfaisant pour la ou les variables d'intérêt et assurer un échantillon spatialement équilibré, tout en n'augmentant pas indûment la quantité globale d'unités d'échantillonnage**. Afin de remédier à la non-réponse potentielle, davantage d'unités d'échantillonnage devraient être envisagées.

Le logiciel et scripts R et les feuillets Excel sont utilisés pour faire le calcul.

En cas d'intensification, les analystes/techniciens du laboratoire, les responsables du Laboratoire et de la Méthodologie doivent documenter la relation entre les échantillons intensifiés et les unités

d'échantillonnage existantes. La justification doit être documentée et stockée conformément au modèle 1 dossier MRV du Programme.

Le nombre d'échantillons sera identifié à l'issue de cette étape.

Pour cela, il faut tout d'abord calculer les superficies par strate (décrit dans 2.4.1).

2.4.1. CALCUL DES SUPERFICIES PAR STRATE

- On commence par la préparation des dossiers et des fichiers, mettre le fichier d'UOT (dans le cas ici **uot_2017_2018.tif**) dans le répertoire **input**. Créer le répertoire **output**.
- Dans le logiciel R, copier-coller et éditer le code suivant. Les parties en rouge doivent être définies par l'utilisateur.
- Les fichiers doivent être dans le dossier « working directory » et les fichiers en entrées (la carte de stratification) dans le répertoire input, puis vérifier qu'il y a bien un répertoire output
- Le résultat se trouve dans un fichier .csv dans le dossier output

```
#####
```

```
# La carte de stratification contient  
  
# les classes a utiliser pour l'echantillonnage stratifie  
  
# les utilisation du sol sont :  
  
#----les classes existantes-----  
  
# - foret persistant code (FF)      11  
  
# - deforestation code (FN)        12  
  
# - gain code (NF)                  21  
  
# - non foret persistant code (NN)  22  
  
# - eau persistante (WW)            44  
  
#-----les classes a generer-----  
  
# - buffer sur deforestation (BD)   66  
  
# - buffer sur gain (BG)            88  
  
# on commence par importer les modules  
  
library(raster)
```

```

library(rgdal)

library(spatialEco)

library(xlsx)

# definir le repertoire de travail

setwd("E:/sops/sop1")

#definir les parametres

res <- 30

carte_strat <- "./output/carte_strates.tif"

area_uot <- "./output/area_uot.csv"

nom_classe <- c("FF","FN","NF","NN","WW","BD","BG")

#en entree on a la carte de changement d'UOT

# il doit etre dans le repertoire input

carte_UOT <- raster("./input/uot_2017_2018.tif")

#carte_UOT <- raster("./input/uot_test_mini.tif")

pal = colorRampPalette(c("green","red","magenta","grey","blue"))

plot(carte_UOT,breaks = c(11,12,13,21,22,44),col = pal(5))

#####

## ON ne touche a rien a partir d'ici

## sauf si vous savez ce que vous faites

#####

```

```

# on isole les deforestation du code 12

# on voudrait donc transformer en NA tous les autres pixels

#FN <- subs(carte_UOT,data.frame(id=c(11,22,21,44),v=c(NA,NA,NA,NA)))
FN <- subs(carte_UOT,data.frame(id=c(11,12,22,21,44),v=c(NA,12,NA,NA,NA)))

plot(FN,breaks = c(0,11,12,13,21,22,44),col=pal(5))

BD <- buffer(FN,100)

BD <- calc(BD, fun=function(x){ x[x >0] <- 66; return(x)} )

#on isole aussi les regeneration du code 21

NF <- subs(carte_UOT,data.frame(id=c(11,12,21,22,44),v=c(NA,NA,21,NA,NA)))

plot(NF,breaks = c(0,11,12,13,21,22,44),col=pal(5))

BG <- buffer(NF,100)

BG <- calc(BG, fun=function(x){ x[x >0] <- 88; return(x)} )

plot(BG,breaks = c(0,6,88,90,92,95),col=pal(5))

#on va maintenant combiner les deux buffers

#BD prendra priorite

#la fonction cover remplace les NA dans un raster par les valeurs d'un autre raster

# remplace les NA dans le premier raster avec les valeurs du 2e raster

BDBG <- cover(BD,BG)

```

```

plot(BDBG,breaks = c(0,66,88,99),col=pal(5))

#combiner les valeurs de UOT avec les buffers

#on va remplacer les valeurs de UOT la ou on a BDBG

#l'objectif est d'avoir une carte pour placer les echantillons

#de facon a ce que les echantillons dans les Buffer

#ne se superpose pas avec les echantillons sans FF et NN

#combinons d'abord les buffers avec la FN

#pour eviter que les FN soit ecrasee par les buffers

#de meme pour la NF

bufferFN <- cover(FN,BDBG)

buffer_FN_NF <- cover(NF,bufferFN)

uot_buffer <- cover(buffer_FN_NF,carte_UOT)

#calcul des surfaces

#les uot a partir de la carte uot, non la stratif

nbpix_uot <- as.data.frame(freq(carte_UOT))

nbpix_uot <- subset(nbpix_uot,nbpix_uot$value!="NA")

nbpix_buffer <- as.data.frame(freq(uot_buffer))

nbpix_buffer <- subset(nbpix_buffer,nbpix_buffer$value==66|nbpix_buffer$value==88)

surface_uot <- rbind(nbpix_uot,nbpix_buffer)

surface_uot$surf <- surface_uot$count*res*res/10000

```

```

surface_uot <- cbind(surface_uot,nom_classe)

#exporter les resultats

#possible datatype for raster : LOG1S,INT1S,INT1U,INT2S,INT2U,INT4S,INT4U,FLT4S,FLT8S

write.csv(surface_uot,area_uot)

writeRaster(uot_buffer,filename = carte_strat,"GTiff",datatype ="INT1U",overwrite = TRUE)

```

A la fin du processus, on aura les fichiers .tif de la carte pour l'échantillonnage et un fichier uot_area.csv contenant alors la surface de chaque utilisation du sol. Le fichier uot_area.csv doit ressembler au tableau suivant :

	value	count	surf	nom_classe
1	11	267765	24098.85	FF
2	12	1473	132.57	FN
3	21	4963	446.67	NF
4	22	1143000	102870	NN
5	44	18636	1677.24	WW
6	66	26206	2358.54	BD
7	88	84068	7566.12	BG

On va utiliser le tableau pour la suite sur le calcul du nombre d'échantillons par strate

```

#####PREMIERES ETAPES#####

# ---Clear the workspace

rm(list=ls())

#Load libraries

library(raster)

library(rgdal)

library(stringr)

```

```

#####DEFINIR DES ENTREES#####

#set working directory. Il faut definir ici le repertoire. Il faut mettre / au lieu de \

wd <- "C:/Dropbox/MADA ERP/CAZ/"

setwd(wd)

# read a raster, GeoTiff or something. Il faut mettre le nom du raster de stratification. Il faut que le raster
soit dans le working directory

Raster <- raster('STR_CAZ.tif')

##Resolution de la carte utilise

Res <- 90

##Fichier avec les superficies par estrate

Output <- "MADA_ERPD_CAZ_Stratification.csv"

###NE TOUCHER CE QUI EST ECRIT CI
DESSUS#####

#####CALCUL DU NOMBRE DE PIXELS PER CLASSE DE LA
CARTE#####

t <- Raster@data@attributes

map_area <- as.data.frame(t)

map_area$Ha

map_area$Ha <- map_area$Count*Res*Res/10000

write.csv(map_area, file= Output, row.names= FALSE)

```

2.4.2. CALCUL DE LA TAILLE DE L'ÉCHANTILLONNAGE

On introduit les superficies par strate dans la feuille de calcul Microsoft Office Excel destiné à calculer le nombre d'unités d'échantillonnage ou taille d'échantillon par strate.

2.5. Étape 4 : Sélection des unités d'échantillonnage

Les analystes/techniciens de Laboratoire en coordination avec les responsables du Laboratoire et de la Méthodologie établissent les emplacements spatiaux des unités d'échantillonnage.

Le logiciel et scripts R sont utilisés pour cela.

Tout d'abord il faut créer un fichier .csv avec deux colonnes, une colonne qui s'appelle "map_value" avec la valeur de la strate et une autre colonne "final" avec le nombre d'échantillons par strate. Il faut mettre le fichier dans le dossier de travail.

Après il faut copier, coller et éditer en R le code suivant. Les parties en rouge doivent être définies par l'utilisateur.

Les fichiers doivent être dans le dossier « working directory ». Le résultat se trouve dans un fichier .csv dans le dossier de travail.

La sortie est un SHP avec les échantillons ET un fichier .CSV qu'on pourra afficher sur Collect Earth.

Important : Confirmer que les points sont dans le centroïde des pixels de la carte de stratification.

Important : Il faut nommer les archives de sortie d'une façon qui permette l'organisation et l'archivage de l'information.

Important : Chaque unité d'échantillonnage doit avoir un code unique.

```
#####PREMIERES ETAPES#####  
  
# ---Clear the workspace  
  
rm(list=ls())  
  
#Load libraries  
  
library(raster)  
  
library(rgdal)
```

```

library(stringr)

#####DEFINIR DES ENTREES#####

#set working directory. Il faut definir ici le repertoire. Il faut mettre / au lieu de \

wd <- "C:/Dropbox/MADA ERP/CAZ/"

setwd(wd)

#Set the project code

Proj <- "MAK2"

# read a raster, GeoTiff or something. Il faut mettre le nom du raster de stratification. Il faut que le raster
soit dans le working directory

Raster <- raster('STR_CAZ.tif')

##Resolution de la carte utilise

Res <- 90

#File name with number of samples per stratum. Fichier avec une colonne "map_value" avec la valeur de
la strate et une autre

##colonne "final" avec le nombre dechantillons

rp <- read.csv('Sampling.csv', header = T)

# Fichier de sortie avec les echantillons

output_file <- "test_plots.csv"

# The ISO country code. Code ISO du pays

```

```

countrycode <- 'MDG'

##Nom du SHP avec les points qu'on va creer

Shp <- "Points_echantillonnage"

###NE TOUCHER CE QUI EST ECRIT CI
DESSUS#####

###NE TOUCHER CE QUI EST ECRIT CI
DESSUS#####

#####CALCUL DU NOMBRE DE PIXELS PER CLASSE DE LA
CARTE#####

t <- Raster@data@attributes

map_area <- as.data.frame(t)

map_area$Ha

map_area$Ha <- map_area$Count*Res*Res/10000

#We paste the map_area in the rp

rp$map_area <- map_area$Count

#####SELECTION DES
ECHANTILLONS#####3

#beginCluster()

##### Generate 10x times the number of points from overall sample

```

```

rand_sample <- data.frame(sampleRandom(Raster,(sum(rp$final) *10 +
log((sum(rp$map_area))))),xy=TRUE)

names(rand_sample) <- c("x_coord","y_coord","map_value")

rand_sample$id <- row(rand_sample)[,1]

rp2 <- merge(rp,data.frame(table(rand_sample$map_value)),by.x="map_value",by.y="Var1",all.x=T)
rp2[is.na(rp2)]<-0

##### Create the list of classes that need to be specifically sampled
to_rtp <- rp2[rp2$Freq < rp2$final,]$map_value

##### Create the list of classes that are enough represented in the random sampling
to_spl <- rp2[rp2$Freq >= rp2$final,]$map_value

##### Sample points from the first class
i = 1

final <- rand_sample[ rand_sample$id %in%
  sample(rand_sample[rand_sample$map_value %in% c(to_spl[i],to_rtp[i]),]$id,
  rp2[rp2$map_value %in% c(to_spl[i],to_rtp[i]),]$final),]

##### Loop into the well represented classes, sample and append
if(length(to_spl) > 1){

```

```

for(i in 2:length(to_spl)){

tmp <- rand_sample[ rand_sample$Sid %in%

      sample(rand_sample[rand_sample$map_value == to_spl[i,]$Sid, rp2[rp2$map_value ==
to_spl[i,]$final),]

      final <- rbind(final,tmp)

}

}

##### Loop into the subrepresented classes, raster_to_point then append

if(length(to_rtp) > 0){

for(i in 1:length(to_rtp)){

tmp_rtp <- as.data.frame(rasterToPoints(Raster,fun=function(rast){rast==to_rtp[i]}))

names(tmp_rtp) <- c("x_coord","y_coord","map_value")

tmp_rtp$Sid<-row(tmp_rtp)[,1]

sampling <- min(rp2[rp2$map_value == to_rtp[i,]$final,

      rp2[rp2$map_value == to_rtp[i,]$map_area)

tmp<-tmp_rtp[tmp_rtp$Sid

      %in%

      sample(tmp_rtp[tmp_rtp$map_value == to_rtp[i,]$Sid,

      sampling

      ),

      ]

final <- rbind(final,tmp)

}

```

```

}

#endCluster()

points <- final

#####
#####

##### Create vector layer with the points

sp_df<-SpatialPointsDataFrame(
  coords=points[,c(1,2)],
  data=data.frame(points[,c(3)]),
  proj4string=CRS(proj4string(Raster))
)

sp_df <- spTransform(sp_df,CRS("+proj=longlat +datum=WGS84"))

#####ON VA CREER LE CSV#####3

coord <- sp_df@coords
coord.sp <- SpatialPoints(coord)
coord.df <- as.data.frame(coord)
coord.spdf <- SpatialPointsDataFrame(coord.sp, coord.df)

#download province boundaries for the country

```

```

adm <- getData ('GADM', country= countrycode, level=1)

#match the coordinate systems for the sample points and the boundaries

proj4string(coord.spdf) <-proj4string(adm)

adm1 <- over(coord.spdf, adm)

nsamples <- nrow(coord)

id <- matrix(sample(1:nsamples , nsamples , replace=F),nrow = nsamples , ncol =1, dimnames=
list(NULL,c("ID")))

id <- as.data.frame(id)

id$ID <- paste(Proj,"-",id$ID)

YCoordinate <- coord[,2]

XCoordinate <- coord[,1]

elevation <- getData("alt", country = countrycode)

slope <- terrain(elevation, opt = "slope")

aspect <- terrain(elevation, opt = "aspect")

elevation <- extract(elevation, cbind(coord[,1], coord[,2]))

slope <- extract(slope, cbind(coord[,1], coord[,2]))

aspect <- extract(aspect, cbind(coord[,1], coord[,2]))

map_code <- sp_df@data[, 1]

#write CSV file, this can be used directly in Collect Earth

m <- data.frame(id, YCoordinate, XCoordinate, elevation, slope, aspect, map_code) #, cadrillage)

write.csv(m, file= output_file, row.names= FALSE)

```

```
# write it out to a shapefile for further processing
```

```
writeOGR(obj = sp_df, dsn = Shp, layer = Shp, driver="ESRI Shapefile") # this is in geographical projection
```

2.6. Important: Revue de l'échantillonnage et finalisation de la stratification

Dans notre cas, en ce qui concerne ce SOP1 qui explique l'échantillonnage, une re-stratification pourrait être faite lorsque l'on a les données de l'échantillonnage pilote. Ces données seraient analysées et dans le cas d'erreurs d'omission dans les strates stables, une exploration qualitative pourra être faite pour évaluer la cause des erreurs.

On peut ensuite alors éventuellement décider de faire la re-stratification en divisant la strate stable en plusieurs strates, pour réduire ainsi la taille de la strate et augmenter le nombre d'échantillons émis par strate.

La génération des nouveaux échantillons se fait de la même manière: en suivant les étapes décrites dans ce SOP en respectant la nouvelle stratification.

ANNEXES

ANNEXE 1: Modèle 1

FORM 1: SAMPLING DESIGN	
Purpose and scope	<i>Provide a short description of the purpose and the scope for which this sampling design is created (for example sampling design for the national FREL of country X)</i>
Version	<i>Insert version number of the sampling design</i>
Date	<i>Insert date on which this version of the sampling design was documented</i>

Basic characteristics of the sampling design

Type of sampling and sample units	<i>Document the selected type of sampling (SRS / SYS /STR) and type of sampling units (map pixels / points / others)</i>
Shape and size of the spatial support	<i>Document the selected shape and size of the spatial support to be used by the interpreter to assign labels to the sample units</i>
Explanation	<p><i>Document the main considerations for choosing the selected options so in the future other can understand the choices made</i></p> <p><i>If the selected sample design is different from previous sample designs used in the RL or previous monitoring periods, the documentation shall also address how the change in design affects the comparability and accuracy of the results.</i></p> <p><i>If a pilot survey has been conducted as noted in the Instructions, explain how the lessons learned from this pilot survey have served to inform the stratification criteria</i></p>

Definitions of strata (for cases where stratified [random/ systematic] sampling has been selected)

Stratum number / Code	Stratum name	Description of the stratum	Area in stratification map a_h
	<i>Include here</i>		a_h
	<i>Include here</i>		a_h
	<i>Include here</i>		a_h
			A

Explanation:

Document the main considerations for choosing the selected stratification scheme so in the future other can understand the choices made.

If a pilot survey has been conducted as noted in the Instructions, explain how the lessons learned from this pilot survey have served to inform the stratification criteria.

If post-stratification has been applied (e.g. a stable forest class stratum has been improved and divided in a stable forest class post-stratum and a deforestation 2 post-stratum), explain the criteria and the process for defining the post-strata.

Number of sample units allocated

Include one table per variable of interest or key variable of interest (i.e. area of deforestation in the period of analysis, area of forest degradation in the period of analysis).

Stratum name	Number of sample units n_h	Expected proportion p_h	Stratum weight W_h	Expected standard error $S(p_h)$	Expected percentage uncertainty of the proportion $U\%(p_h)$
<i>Include here</i>	n_h	p_h	W_h	$\sqrt{p_h(1-p_h)/(n_h-1)}$	$t_{a,df} * S(p_h) / p_h$
<i>Include here</i>	n_h	p_h	W_h	$\sqrt{p_h(1-p_h)/(n_h-1)}$	$t_{a,df} * S(p_h) / p_h$
<i>Include here</i>	n_h	p_h	W_h	$\sqrt{p_h(1-p_h)/(n_h-1)}$	$t_{a,df} * S(p_h) / p_h$
	N	$\sum_{h=1}^H W_h * p_h$		$\sum_{h=1}^H W_h^2 * S^2(p_h)$	$U\%(p) = \frac{t_{a,df}}{p} * \sum_{h=1}^H W_h^2 * S^2(p_h)$

Explanation:

Document the main considerations for allocating the sample units to strata.

If a pilot survey has been conducted as noted in the Instructions, explain how the lessons learned from this pilot survey have served to inform the stratification criteria.

If intensification on a specific stratum has been applied, document how the intensified samples relate to the existing sample units, e.g. in systematic design, are the intensified samples aligned with the existing samples?

Sample unit allocation

Document the main steps taken to allocate the sample units to strata or the region of interest, including the steps for randomizing the location.