Instituto Nacional de Conservación y Desarrollo Forestal, Áreas Protegidas y Vida Silvestre (ICF)

Servicio Forestal de Los Estados Unidos de Norteamérica Programas Internacionales

Estimates Methods for the National Forest Inventory of Honduras

Preparado por:	*Presentado a:*
Charles T. "Chip" Scott Consultor	Betina Salgado, ICF Craig Wayson, USDAFS-IP

Tegucigalpa, Honduras
Diciembre, 2019

Introduction

This document details the estimation procedures for the National Forest Inventory (NFI) of Honduras. These methods were applied in a companion document, *Honduras IFN Ciclo 3 Estimation 2019-12-03.xlsm*, which applies the procedures using Excel to the data from the third cycle of the NFI.

Background

Honduras is in the middle of the third cycle of the National Forest Inventory. The first NFI was completed in 2006. It was conducted by the UN Food and Agriculture Organization (FAO) for COHDEFOR (precursor to ICF). The second was financed by the European Union through the MOSEF and FORCUENCAS projects. While it was built upon the initial survey, there were several enhancements. For the third NFI, ICF received technical assistance from International Programs of the U.S. Forest Service on the design. ICF has completed 3 of the 5 years of data collection for the third cycle. Due to the design, this subset of plots is representative of the entire country, thus can be used for estimation.

Sampling Design Overview

While the sampling designs have been documented elsewhere, the following provides a brief overview of the design at each of the three occasions (measurement cycles).

Design at Cycle 1

In 2005, FAO placed a grid of roughly 200 plots across the country. 181 were visited and measured. The remaining ones were in inaccessible areas of the country. FAO used the plot design that was common for them at that time. The largest trees were measured on four 20x250m subplots arranged in a square circumscribing 1 km^2. The emphasis was on the forest resource and not on carbon.

Design at Cycle 2

In 2011 and 2015, ICF remeasured 168 of the original 181 plots. They established 95 new plots to improve the precision. To improve efficiency, they reduced the size of the subplots from 20x250 m to 20x130 m.

Design at Cycle 3

The designed changed yet again for the third cycle. The plot size was reduced, and the sample size was doubled. The plot design was changed to three circular plots overlaid on the first 20x150 m subplot from cycle 2. This was done in part because large rectangular plots are more difficult to remeasure on a tree-by-tree basis. Circular plots with a known distance and azimuth to each tree are much easier to remeasure, thus allowing for the estimation of the components of change: ingrowth, accretion, mortality and harvest. With the reduction in plot

area, the sample size was doubled, with a target when done of about 420 plots. So far in the third cycle, 228 plots have been measured in the first 3 of 5 annual panels. An additional 22 plots were inaccessible. Of those measured, 125 were remeasurements of cycle 2 plots, thus roughly half the plots are remeasured, and half the plots are new.

Due to the keen interest in the amount of carbon in mangrove forests, an intensification was conducted in four different strata within the mangrove portion of Honduras. 54 plots were established in these four parts of the country, each with a different sampling intensity. The plots were distributed in a spatially balanced way within each stratum using a Space Filling Curve approach (Lister and Scott 2009). The rest of the country (upland) was treated as stratum 5, which is also a spatially balanced sample. The plots were systematically assigned to five annual panels so that 20% of the plots are to be measured each year and so that the subset is a spatially balanced sample of the country.

Stratum	Area (ha)	Sample Size	Ha/plot
1	2,463	14	175.9
2	5,324	16	332.7
3	4,028	14	287.7
4	40,156	10	4,015.6
5	11,197,229	228	49110.7
Total	11,249,200	282	39890.8

Thus, the total number of plots used for estimation at this time is: 228 + 54 = 282. Soil samples were collected in the mangroves, but not yet in the upland areas.

Estimation Methods Overview

This section describes the estimation steps in detail and includes the statistical formulas. Estimation starts with the observation then works in a hierarchal way to the plot, to the stratum, and to the final estimate. The flow is as follows:

- Start with the observation on sampled element. Some attributes of interest are calculated values from models, such as for basal area, volume, biomass, etc.
- Sum the values of the attribute of interest for all observations on the plot within the domain of interest.
- Compute the area measured on the plot (accounting for inaccessible portions) and the area within the domain of interest.
- Average across the stratum while adjusting for partial plots
- Compute the variance based on the sampling design
- Repeat for each domain of interest
- Compute at least once each for all plots (inaccessible and sampled) for comparison to the results using only sampled plots.

A hierarchy is evident from the above. The highest level is the population, in this case, all land and water in the country of Honduras. Next, the population is divided into five strata at cycle 3; only one stratum was used at cycles 1 and 2. Within each stratum a set of plots was selected. Within each plot, several subplot kinds and sizes are used to sample the different resource characteristics, such as trees, dead and down material, and soils. The notation used below reflects this hierarchy.

- Population – the entire area of interest, in this case, the area within the boundaries of Honduras, including both land and inland water.
- Strata – mutually exclusive subdivisions of the population. At cycles 1 and 2, there was only on stratum. At cycle 3, four mangrove areas were identified and sampled at different intensities. The 5th stratum was all other areas, mostly upland.
- Domains – classifications of individual plots or sampling elements within plots (Scott et al 2005), such as land use class, tree species or diameter class. These are observations and not constructs of the sampling design (unlike strata). Domains are used in the estimation process to group the estimates, such as for making a table of biomass by species or diameter class.

Computing Plot-level Attributes by Domain

The first step is to specify the attribute of interest and how it was observed. For example, tree basal area is observed by measuring the diameter of each sample tree, then computing from the diameter the cross-sectional area of the tree. Other tree attributes of interest include the number of trees per hectare, stem volume, and biomass. Much like basal area, stem volume and biomass are not observed directly. Volume was already computed using a volume model based on the species group, the diameter, and the commercial height. Biomass was already computed similarly with a biomass model. For details on how this was done, see *Procedimientos de calculos INF Honduras 1ra y 2da medicion.docx* and *Procedimientos de calculos INF Honduras 3ra medicion.docx*.

For the plot design differed by cycle. The table below shows the area sampled for each of the subplot kinds.

Subplot Kind	Cycle 1	Cycle 2	Cycle 3
Trees >20 cm in diameter	2.0 ha	1.04 ha	0.21206 ha
Trees 10-20 cm in diameter	2.0 ha	1.04 ha	0.033929 ha
Trees 1.3 m in ht to <10 cm in diameter	0.24 ha	0.16 ha	0.0084823 ha
Necromass >=10 cm	none		63 m
Necromass <10 cm	none		15 m
Soil sample	none	none	Strata 1-4 only

At cycles 1 and 2, the proportion of the plot by land class was determined by determining the proportion of the center line by land class. At cycle 3, there are 5 points in each of the 3 subplot locations. Each point is assigned to a single land class, so the proportion is computed using these 15 points. Each tree is automatically assigned to the closest point's land class.

The next consideration is whether the attribute is in the domain of interest. For example, the domain of interest might be a diameter class. Only trees in the diameter class of interest would be added up in the plot total. This would be repeated for each class of the domain of interest plus the total across all classes. The result is called a one-way table: an estimate for each class of the domain of interest and an estimate for the total.

Notation

Before introducing the estimation formulas, the notation used is presented. The attribute of interest is *y*, such as biomass. For ratio estimates, the denominator, *x*, is often the area of interest, such as forest area or a specific forest type. The area sampled on a plot is *a*. The letter *t* is used as an indicator variable (either 0 or 1) for whether the observation is in the domain of interest or not, such as a tree being in a specific diameter class or forest type. The domain of interest is designated as *d*.   Subscripts are used to specify characteristics of the attributes (*y, x, a* and *t*), mostly to indicate the hierarchy of sampling from trees to plots to strata to the population and over time.

> g = population – a geographic subdivision of the entire population. In this case, there is only one subpopulation – the entire population.
>
> h = stratum – each subpopulation can be divided into strata with known proportions of the total subpopulation area. Stratification is used to account for the different sampling intensities at cycle 3.
>
> i = plot – refers to the collection of all subplots which are established at a GPS location – the sampling unit
>
> j = subplot, such as for the different sized trees
>
> k = observational unit, such as a tree

Thus, the biomass of tree k on subplot j of plot i in stratum h in subpopulation g is denoted as: y_{ghijk}.

The attributes need to be expressed at the plot level. Trees are often sampled in nested plots, as was noted above. In order to be able to combine estimates across the different subplot sizes, the approach here is to put them all on the same basis as the largest subplot. To do this, the observations must be summed across the plot and then expanded to the size of the largest subplot, 2.0, 1.04, or 0.21206 ha in this case. The expansion factor for a particular subplot size and its observational units is simply the area of the biggest plot divided by the area used to sample the observation unit (e.g., tree). This can be computed as:

$$y_{ghid} = \Sigma_j^J \left(y_{ghijd} \frac{\Sigma_i^{n_{gh}} a_{ghiJ}}{\Sigma_i^{n_{gh}} a_{ghij}} \right) = \Sigma_j^J \left(\frac{\Sigma_i^{n_{gh}} a_{ghiJ}}{\Sigma_i^{n_{gh}} a_{ghij}} \right) \Sigma_k \, y_{ghijk} t_{ghijkd} \qquad \text{(a)}$$

where:

> y_{ghid} = sum of the attribute of interest in domain of interest d on plot i in stratum h in subpopulation g
>
> y_{ghijd} = sum of the attribute of interest in domain of interest d in subpopulation g in stratum h on plot i on subplot size j, such as the sum of sampled tree volumes. Note that observation unit values are simply summed and are <u>not</u> expressed on a per unit area basis.
>
> a_{ghij} = subplot size j area measured on plot i in stratum h in subpopulation g.
>
> y_{ghijk} = the attribute of interest of observation unit k on subplot size j in plot i in stratum h in subpopulation g.

t_{ghijkd} = indicator variable which is 1 if observation unit k on subplot size j on plot i in stratum h in subpopulation g is in the domain of interest, such as a specific forest type; 0 otherwise.

a_{ghiJ} = area of the largest subplot size, J, measured on plot i in stratum h in subpopulation g. In this case, it is 1,000 m^2.

The separate adjustment factor is computed for each subplot size. The factor is computed as the total area measured on the largest subplot divided by the total area measured for each of the subplots for all plots in the stratum. For trees on the largest subplot, the factor is 1. The factor for trees on smaller subplots, the factor is greater than 1 and is the same for all trees sampled on that size across the stratum. The adjustment is NOT made at the plot level, since that would lead to a biased result. The reason for expanding to the area of the largest plot has to do with the next step.

Ratio-to-Size Estimators

Ratio estimators come in many forms. In this document, three forms are used. The simplest is the estimation of proportions, such as the proportion of the country that is forested. While many people think this is to be estimated from maps, it can accurately (perhaps with less bias) be estimated as the average of the proportion of each plot that is forested, then weighted by strata areas. In the special case where all plots are completely forested or completely nonforest, then the proportion above is the same as the ratio of the number of forested plots to the total number of plots which is an accurate estimate of the proportion of forest in the country. Statistically, this is just a simple mean rather than a ratio estimate since the denominator (the number of plots) is not a random variable.

The second type of ratio estimate is often referred to as a ratio of means, where one mean is divided by another to form the ratio estimate. For example, the mean forest biomass/ha of land can be estimated across all plots and the mean forest proportion can be estimated as above. The ratio of these two yields the biomass per ha of forest. It is a good way of estimating the mean per ha within any given class (domain).

The Ratio-to-Size estimator is a special type of ratio estimator that is used when the cluster size of a two-stage sample varies, thus taking advantage of the relationship between the total of the attribute observed and the size of the sampling unit (plot or cluster). That is, the larger the plot area, you would expect a larger the sum of the attribute observed. This estimator is used here to account for different plot sizes resulting from portions of them that may be inaccessible. In the Simple Random Sampling case, the estimator can also be used to take advantage of the relationship between any two attributes measured on the plot.

As used here, the Ratio-to-Size estimator will be applied to each stratum solely to account for missing (inaccessible) portions of plots. Both the attribute of interest and the area on which it was sampled need to be tracked. Any missing portions of plots are assumed to be like the average of the rest of the plots in the stratum, which is better than assuming they are like the average across all strata. This approach has been shown to be less biased than making the

correction at the plot level (Bechtold and Scott 2005). The estimator for the stratum mean of the sampled area in the domain of interest is:

$$\bar{X}_{ghd'} = \frac{\Sigma_i x_{ghid'}}{\Sigma_i a_{ghi}} = \frac{\Sigma_i \Sigma_k t_{ghikd'} a_{ghik}}{\Sigma_i \Sigma_k a_{ghik}} \tag{b}$$

where:

$x_{ghid'}$ = area of plot i in domain of interest d' in stratum h in subpopulation g. While domains d and d' can differ, they are typically either the same or d is a subset of d', such as d being trees greater than 10 cm in diameter on forest land and d' being forest land.

a_{ghik} = area of plot i in stratum h in subpopulation g that is in condition k, such as forest or other land class. If the entire plot is in a single condition, the area is equal to a (the plot area). The area of condition k is estimated based on either the proportion of points within the plot, such that a_{ghik} = a*(number of points in k)/(total number of points on plot i), or a*(the proportion of a line through the plot that is in k). If the condition is not measured due to being out of population or inaccessible, then the area is 0 (unless the domain being estimated is the inaccessible land class).

a_{ghi} = area of plot i in stratum h in subpopulation g across all domains of interest. If this includes inaccessible or out-of-population portions, then a_{ghi} = a.

$t_{ghikd'}$ = indicator variable which is 1 if condition k on plot i in stratum h in subpopulation g is in the domain of interest d', such as a specific forest type; 0 otherwise.

The stratum variance of the area in the domain is:

$$s^2_{xghd'} = \frac{n^2_{gh}}{n_{gh}-1} \frac{\Sigma_i^{n_{gh}} x^2_{ghid'} - 2\bar{X}_{ghd'} \Sigma_i^{n_{gh}} x_{ghid'} a_{ghi} + \bar{X}^2_{ghd'} \Sigma_i^{n_h} a^2_{ghi}}{\left(\Sigma_i^{n_{gh}} a_{ghi}\right)^2} \tag{c}$$

The estimator for the stratum mean of the attribute in the domain of interest is:

$$\bar{Y}_{ghd} = \frac{\Sigma_i y_{ghid}}{\Sigma_i a_{ghi}} \tag{d}$$

where:

y_{ghid} = sum of the attribute of interest in domain of interest d on plot i in stratum h in subpopulation g. It is computed using (a) above.

The stratum variance of the attribute of interest in the domain is:

$$s^2_{yghd} = \frac{n^2_{gh}}{n_{gh}-1} \frac{\Sigma_i^{n_{gh}} y^2_{ghid} - 2\bar{Y}_{ghd} \Sigma_i^{n_{gh}} y_{ghid} a_{ghi} + \bar{Y}^2_{ghd} \Sigma_i^{n_{gh}} a^2_{ghi}}{\left(\Sigma_i^{n_{gh}} a_{ghi}\right)^2} \tag{e}$$

In preparation for ratio-of-means estimation, the stratum covariance is:

$$cov(Y_{ghd}, X_{ghd'}) = \frac{n_{gh}^2}{(n_{gh}-1)} \frac{\sum_i^{n_{gh}}(y_{ghid}-a_{ghi}\bar{y}_{ghd})(x_{ghid'}-a_{ghi}\bar{x}_{ghd'})}{\left(\sum_i^{n_{gh}} a_{ghi}\right)^2}$$

$$= \frac{n_{gh}^2}{(n_{gh}-1)} \frac{\sum_i^{n_{gh}} y_{ghid}x_{ghid'} - \bar{y}_{ghd}\sum_i^{n_{gh}} a_{ghi}x_{ghid'} - \bar{x}_{ghd}\sum_i^{n_{gh}} a_{ghi}y_{ghid'} + \left(\sum_i^{n_h} a_{hi}^2\right)\bar{y}_{ghd}\bar{x}_{ghd'}}{\left(\sum_i^{n_h} a_{ghi}\right)^2} \qquad \text{(f)}$$

While the first line of (f) can be recognized as the crossproduct of the residuals, the second line is easier computationally.

To this point, we have assumed Simple Random Sampling within each stratum. However, rather than using the usual mean and variance estimators, we have used the Ratio-to-Size estimator of the mean and variance. In the case where every plot is fully measured, then the Ratio-to-Size mean and variance are exactly equal to the usual Simple Random Sampling estimates.

Stratified Random Sampling Estimators

As was noted above, Cycles 1 and 2 used a single stratum, thus only Simple Random Sampling estimators are required. Cycle 3 uses 5 strata. The following Stratified Randomestimators apply to both situations with the number of strata, H, equaling 1 in the first case and 5 in the second. The estimate of the average area across all strata is:

$$\bar{X}_{gd'} = \sum_h^H W_{gh}\bar{X}_{ghd'} \qquad \text{(g)}$$

where:
    W_{gh} = stratum h weight in subpopulation g which is the proportion of the grid points falling in the stratum. For a single stratum, $W_{gh} = 1$.

        $= \frac{A_{gh}}{A_g}$

    A_{gh} = the area falling in the stratum h in subpopulation g

    A_g = the total area in subpopulation g

Its variance is:

$$v(\bar{X}_{gd'}) = \sum_h^H W_{gh}^2 \frac{s_{xghd'}^2}{n_{gh}} \qquad \text{(h)}$$

Similarly, the mean of the attribute of interest and its variance are:

$$\bar{Y}_{gd} = \sum_h^H W_{gh}\bar{Y}_{ghd} \qquad \text{(i)}$$

$$v(\bar{Y}_{gd}) = \sum_h^H W_{gh}^2 \frac{s_{ygh}^2}{n_{gh}} \qquad \text{(j)}$$

In preparation for ratio-of-means estimation, the covariance is:

$$cov\left(\bar{Y}_{gd}, \bar{X}_{gd'}\right) = \sum_h^H W_{gh}^2 \frac{cov(Y_{ghd}, X_{ghd'})}{n_{gh}}$$

(k)

where the stratum covariance was given in (f).

If there is more than one subpopulation, the population means are the weighted sum of the subpopulation means:

$$\bar{X}_d = \sum_g^G \frac{A_g}{A} \bar{X}_{gd}$$

(l)

$$\bar{Y}_d = \sum_g^G \frac{A_g}{A} \bar{Y}_{gd}$$

(m)

Since the subpopulations are independent, then their variances are computed using the weights squared:

$$v(\bar{X}_d) = \sum_g^G \frac{A_g^2}{A^2} v(\bar{X}_{gd})$$

(n)

$$v(\bar{Y}_d) = \sum_g^G \frac{A_g^2}{A^2} v(\bar{Y}_{gd})$$

(o)

Ratio Estimators

The estimator of the ratio between the two means (ratio of means) provides an estimate on a per hectare basis for just the domain of interest, such as the mean biomass per hectare of forest rather than across all lands. It is:

$$\hat{R}_{dd'} = \frac{\bar{Y}_d}{\bar{X}_{d'}} = \frac{\sum_g^G A_g \bar{Y}_{gd}}{\sum_g^G A_g \bar{X}_{gd'}}$$

(p)

Its variance is:

$$v\left(\hat{R}_{dd'}\right) = \frac{1}{\bar{X}_{d'}^2}\left[v(\bar{Y}_d) + \hat{R}_{dd'}^2 \, v(\bar{X}_{d'}) - 2\,\hat{R}_{dd'}\, cov(\bar{Y}_d, \bar{X}_{d'})\right]$$

(q)

The covariance term is:

$$cov(\bar{Y}_d, \bar{X}_{d'}) = \sum_g^G \frac{A_g^2}{A^2} cov(\bar{Y}_{gd}, \bar{X}_{gd'})$$

(r)

This completes the estimation procedures for either the Cycle 1 plots or the new plots.

Estimation with and without Inaccessible Plots

Normally, all plots are assumed to be included in the estimators above. However, it is common for portions of plots or entire plots to be lacking data. The most common reasons are that access was denied by the owner or local people or that it was too dangerous to occupy the location. The partial plot situation was already addressed using the Ratio-to-Size estimator – the sampled area is in the denominator.

In the case where whole plots are missing data, most estimates will be made by simply not including these plots. That is, these estimates are based on sampled plots only. The sample sizes, n, would not include the plots which were entirely inaccessible.

However, it is recommended that at least the area of the non-sampled region be estimated using the data. In this case both the non-sampled plots plus the non-sampled portions of plots would be included. To do this, those plot areas that were missing data would be coded with a land use or land cover class of "Non-sampled" or "Inaccessible". The sample sizes would then include all sampled as well as non-sampled plots. The plot areas, a_{ghi}, would all equal a. The domain variable would be land use/cover class so one of the domains would be the non-sampled class. With these adjustments, the same estimation procedures above (formulas (a) through (r)) would be used.

Estimation of Change

While plots have been remeasured from cycles 1 to 2 and from cycle 2 to 3, to attempt has been made at this time to compute plot level change estimates. This is largely due to the change in plot designs. This will become more important at cycle 4 when individual trees are remeasured and used to compute the components of change: ingrowth (new trees), accretion (tree growth), mortality and harvest. Estimates of change at this time are simply computed as the net difference between estimates from one cycle to the next. A conservative estimate of the variance of the net change is to sum the variances of the two estimates. A more appropriate estimate of the variance of net change accounts for the subset of remeasured plots and their covariance:

$$ v(\bar{Y}_d - \bar{X}_{d\prime}) = v(\bar{Y}_d) + v(\bar{X}_{d\prime}) - \frac{n_{t,t-1}}{n_t n_{t-1}} \, cov(\bar{Y}_d \, , \, \bar{X}_{d\prime}) \qquad\qquad (s) $$

where:
- n_t = total number of plots measured at time t
- n_{t-1} = total number of plots measured at time t-1
- $n_{t,t-1}$ = number of plots remeasured from time t-1 to t

Since the previous cycle is assumed to be the reference point for change, Cycle t-1 sampled areas are to be used for a_{ghi}.

References

Bechtold, William A. and Scott, Charles T. 2005. The Forest Inventory and Analysis plot design. Pages 27-42 *in* William A. Bechtold and Paul L. Patterson, eds. The enhanced Forest Inventory and Analysis program--national sampling design and estimation procedures. U.S. Department of Agriculture, Forest Service, Southern Research Station, General Technical Report SRS-80, Asheville, NC. http://www.srs.fs.usda.gov/pubs/20371

Lister, Andrew J. and Scott, Charles T. 2009. Use of space-filling curves to select sample locations in natural resource monitoring studies. Environmental Monitoring and Assessment 149:71-80. http://treesearch.fs.fed.us/pubs/13672

Scott, Charles T., Bechtold, William A., Reams, Gregory A., Smith, William D., Westfall, James A., Hansen, Mark H., and Moisen, Gretchen G. 2005. Sample-based estimators used by the Forest Inventory and Analysis National Information Management System. Pages 43-67 *in* William A. Bechtold and Paul L. Patterson, eds. The enhanced Forest Inventory and Analysis program--national sampling design and estimation procedures. USDA Forest Service, Southern Research Station, General Technical Report SRS-80, Asheville, NC. http://www.srs.fs.usda.gov/pubs/20371