



Food and Agriculture
Organization of the
United Nations

NFI Design Plan for Indonesian Forests



Project:

Strengthening Indonesian forest and land monitoring for climate actions
(GCP/INS/087/NOR)

FAO, Indonesia
April 2023

NFI Design Plan for Indonesian Forests

Project:
Strengthening Indonesian forest and land monitoring for climate actions
(GCP/INS/087/NOR)

FAO, Indonesia
April 2023

NFI Design Plan for Indonesian Forests

ISBN:

.....

Project:

Strengthening Indonesian forest and land monitoring for climate actions (GCP/INS/087/NOR)

FAO, Indonesia

April 2023

Contents

Executive Summary	vi
Abbreviations	viii
1 Introduction	1
1.1 Background	1
1.2 Purpose of NFI	1
1.3 Basic concept	3
1.4 Objectives	5
2 Proposed Forest Inventory Design	6
2.1 Sampling designs options	6
2.2 Stratification and spatial arrangement of plots	7
2.3 Plot design	9
2.4 Cluster design	11
2.5 Sampling frame	15
2.6 Accessibility of sampling sites	17
2.7 Cost estimates	19
2.8 Biodiversity, trees outside of forest and non-timber forest products	20
3 Development of Final Forest Inventory Design	21
3.1 Refinement of old maps and statistics	21
3.2 Development of biometric models	22
3.3 Development of rapid change monitoring procedure	22
3.4 Practical tests and field surveys	23
3.5 Development of NFI Information System	23
4 Acknowledgements	24
5 References	25
Appendix	27
Appendix 1. Review of alternative data and methods	27
1.1 Sampling strategies	27
1.2 Review of remote sensing and vegetation mapping alternatives	30
1.3 Potential materials for stratification of vegetation in NFI	31
Appendix 2. Forest Structure and NFI Design Analysis	37
2.1 Analysis of Field Plot Data	37
2.2 Description of forest structure and scale in forests	50
2.3 Comparison cluster sampling alternatives using ICC	55
2.4 Sample grids in different islands	58
Appendix 3. CORE and OPTIONAL entities and variables	60
3.1 CORE and OPTIONAL NFI entities and variables	60

List of figures and tables

Figures

1. Focus of NFI is at the national level reporting.	2
2. Islands and distribution of the subset of the past NFI field sample plots used in this NFI re-design analysis. The colours show the proposed 7 island groups for stratifying the new NFI.	4
3. Multiphase sampling can be used to estimate forest attributes with remote sensing data.	6
4. Examples of hexagonal grids overlaid with land use map and satellite image-based forest area (dark green) in Jawa (5 km grid) and Kalimantan (20 km grid).	8
5. Example of land use/forest type sections distribution within a nested circular plot	10
6. Example of regular pattern grid generated to study intracluster correlation and define feasible spatial design.	11
7. Layout of proposed plot design (trees).	13
8. Bamboo, stumps, deadwood, rattan and optional objects to be recorded in plots.	14
9. Illustration of regular grid and hexagonal grids in Sumatra.	17
10. Accessibility analysis based on elevation.	18
11. Distance to roads and rivers in Kalimantan (left). The shortest distance was computed for each cluster's reference point. Red color indicates over 12 km distances (middle) and 5 km distances (right).	18
12. Areas to be covered with Airborne Laser Scanning (ALS) sample lines (5-10 lines per area).	18
13. Flowchart of inventory process.	21
14. Example of canopy height model covering entire Indonesia, 2019. Ground resolution is 30 m.	33
15. Example of canopy height model covering entire Indonesia, 2020. Ground resolution is 10 m.	34
16. Example of GEDI above-ground biomass data.	35
17. ETH Canopy Height image (top left), Indonesian Land cover map (top right), 15 m height threshold forest map (low left) and 20 m height threshold forest map (low right) describe status of land cover.	36
18. Field sample plots in inventory clusters.	37
19. Layout of a Permanent Sample Plot.	38
20. Number of sample plots by year of assessment.	39
21. Demonstrated clustering effect in permanent sample plots within a subset of plots in Kalimantan data.	40
22. DBH distributions by islands (drylands only) and in mangrove forest. The turquoise bars represent the recorded trees in RUs. Dimensions of RU squares are presented as in equal size circles (text boxes).	41
23. Examples of computed DBH distributions with different plot design options for Bali Nusa and Jawa.	43

24. Plot design options for further analysis (drylands and mangroves across all islands).	44
25. Examined plot design options for drylands and mangrove sites with a plot generated from randomly chosen RUs.	45
26. CV of basal area analyzed with different plot and subplot combinations.	46
27. GEDI data (dots) overlaid with canopy height data (left). Variation within forest areas was used in analysis. Also, GEDI AGB was analyzed with land use map categories (right).	51
28. Autocorrelation inside PSPs in Islands.	52
29. Covariance (correlation) of GEDI data (AGB) in inside forest by islands.	53
30. Covariance (correlation) based on ETH canopy height values.	54
31. Regular pattern grid was generated to study intracluster correlation with various cluster designs.	57
32. Regular pattern grids (with varying size of clusters) were generated to study intracluster correlation using tree height data, Distance between plots was varying between 50–500 m.	57
33. Intracluster correlation curves calculated for islands. X-axis is distance between plots and Y-axis is correlation between plots. A plot is equal to a canopy height observation from the ETH data.	58
34. Differences between sparse populations in Jawa (upper part) and dense forest areas in Papua (lower part).	59

Tables

1. Survey unit specification, circular plot design on drylands.	10
2. Survey unit specification, circular plot design for Mangrove stratum	10
3. Comparison of small plot design (plots in tracks without ICC) and larger plot design (3 subplots, correlated) in different Islands.	12
4. Estimated number of samples by islands with preferred cluster plot designs. Sample size for Mangroves is computed for the whole country.	14
5. The proposed sampling unit design compared with 1-ha PSPs	15
6. Proposed spacing when Land Use forest maps are used in stratification. Number of plots in proposed hexagon grids vs original plan (N=Number of Hexagon Plots, t-km ² =1000 km ² in Land Use Map) and distribution of plots in Land Use categories.	16
7. Estimated total NFI costs.	19
8. Estimated additional cost of a NFI project.	20
9. Review of MODIS, Landsat and Sentinel Data.	31
10. Selection of alternative LiDAR datasets.	32
11. Examined concentric circular plot designs.	42
12. Original subset of NFI data: Estimated number of samples on drylands by islands.	48
13. Estimated number of samples on drylands by islands with two methods (25 m radius plot and 3 cluster plots design).	48
14. Estimated number of samples and cumulative area to be sampled by islands with different plot layouts and cluster plot designs.	49
15. Estimated number of samples by islands with preferred cluster plot designs.	50
16. Intracluster correlation of large trees (DBH $\geq$ 40 cm) above-ground biomass calculated for PSP record units.	53
17. GEDI AGB in various Land use/Land cover types in Kalimantan	54
18. GEDI AGB in various Land use/Land cover types in Jawa.	55
19. Alternative plot layouts and Intracluster Correlation Coefficient analyzed with different cluster designs.	56

Executive Summary

This document has been prepared for the project “*Strengthening Indonesian forest and land monitoring for climate actions*” (GCP/INS/087/NOR) in Indonesia. The Government implementing agency is the Ministry of Environment and Forestry (MoEF) and FAO provides technical assistance for the redesign of the National Forest Inventory (NFI) system. The main objective of this document is to provide a proposal for the new NFI design to be applied during the development and implementation of the full-scale National Forest Inventory as a part of the National Forest Monitoring System (NFMS) in Indonesia. The proposed NFI design covers all 5 carbon pools which have to be reported under the UNFCCC. In addition, the proposed sampling design takes into account error estimation and budget constraints.

The proposed NFI design recommends stratifying the country into the following strata: drylands and swamp forests divided into 7 island groups (Sumatra, Jawa, Bali Nusa, Kalimantan, Sulawesi, Maluku, Papua), mangroves, inaccessible sites and trees outside the forests. This plan covers state forests, but a separate plan is under preparation to cover Indonesian customary (private) forest lands.

There will be a different sampling designs for each of these areas, as follows:

- Dryland and swamp forests. Two-phase systematic sampling, with 5–25 km sampling grid depending on Island (Sumatra 17 km, Jawa 5.6 km, Bali Nusa 5.6 km, Kalimantan 17 km, Sulawesi 9.8 km, Maluku 9.8 km, Papua 29 km).
- Mangrove stratum. Includes current mangrove area map(s) and a 200 m buffer zone to capture thin coastal slivers of land caused by digitizing errors. Two-phase sampling design: original grid intensity (of the island) will be intensified to 1 km x 1 km grid for visual interpretation, and in the second phase altogether 30 field plots are allocated to each of the 7 island groups to ensure that the statistical accuracy is likely to be good enough.
- Trees outside forests (canopy cover 10–30%, area 1–6.25 ha). Altogether 30 field plots are allocated to each of the 7 island groups.
- Inaccessible areas (separate ground truthing with detailed remote sensing).

The sampling unit is a plot (sometimes called cluster plot), with 3 concentric circular subplots, and distance between subplots is 100 m on drylands and 50 m on mangroves. All the plots belong into the field sample but those plots that according to prior information (recent very high-resolution remote sensing data) without doubt neither are in forest nor contain trees do not need to be assessed in the field.

We expect that the proposed design produces a sampling error less than 2 % for the forest area estimate and around 2 % for the mean biomass estimate for the whole country. The target sampling error on drylands strata is 5%, and 10% for mangrove forests for the mean biomass estimate.

All field plots are considered permanent for the initial establishment and there are no plans for doing temporary sample points or clusters like the old NFI because that data was not used and took extra time and money. The NFI variables are split into two categories: core variables and optional variables. The core variables will be collected following field procedures that are consistent and uniform across all areas in Indonesia. The optional variables will be collected either by a subset of sample sites or differently by strata. Optional variables are not included into the estimated costs. We recommend collecting socio-economic data in a campaign that is separate from NFI because the effective sampling design for socio-economic data collection is different from the design that is planned for collecting biophysical data.

The NFI cycle in Indonesia is 5 years by the law. The estimated cost of field assessment is USD 7 million, and with Quality Control teams USD 8 million. Based on estimates presented in Chapter 2.3, the total cost of a full scale NFI is in the range of USD 11–12 million (around \$2.4 m/yr). This is considered reasonable by global standards and justified in view of the importance of building confidence in Indonesia's forest data and its potential use for supporting forest carbon trading and results based payments schemes like REDD+.

Abbreviations

AGB	Above Ground Biomass
ALS	Airborne Laser Scanning
asl	Above sea level
BA	Basal area
CV	Coefficient of Variation
DBH	Diameter at the breast height (1.30 m above ground)
DEM	Digital Elevation Model
ETH	Swiss Federal Institute of Technology in Zürich
FAO	Food and Agricultural Organization of the United Nations
FAOR	FAO Representative
FMU	Forest Management Unit
FRA	Forest Resource Assessment
FREL	Forest Reference Emission Level
GEDI	Global Ecosystem Dynamics Investigation
GHG	Green House Gas
GIS	Geographic Information System
GPS	Global Positioning System
ICC	Intracluster Correlation Coefficient
IPCC	Intergovernmental Panel on Climate Change
IPSDH	<i>Direktorat Inventarisasi Dan Pemantauan Sumber Daya Hutan</i> (Directorate of Forest Resources Inventory and Monitoring)
kNN	k-nearest neighbor estimation method
LiDAR	Light Detection and Ranging
LUV	Land Use / Vegetation
MODIS	Moderate Resolution Imaging Spectroradiometer
MoEF	Ministry of Environment and Forestry
NASA	National Aeronautics and Space Administration (USA)
NFI	National Forest Inventory (Inventarisasi Hutan Nasional)
NFMS	National Forest Monitoring System

NTPP	Non-timber forest product
PSP	Permanent Sample Plot
R	R computer program language
REDD	Reducing Emissions from Deforestation and Forest Degradation
REDD+	REDD “plus” refers to “the role of conservation, sustainable management of forests and enhancement of forest carbon stocks.
rl	rl from page 13 formula “Step 3: Calculate Design Effect, $DE=1+(m-1)rl$,
RS	Remote Sensing
RU	Record Unit (i.e. a plot section in NFI permanent sample plot)
SE	Sampling Error
SRS	Simple Random Sampling
TM	Thematic Mapper
TOF	Trees Outside of Forest
TSP	Temporary Sample Plot
UNFCCC	United Nations Framework Convention on Climate Change
VCF	Vegetation Continuous Fields
VHR	Very High Resolution

1 Introduction

1.1 Background

The Government of Indonesia Ministry of Environment and Forestry (MoEF) has asked for FAO technical support to enhance Indonesian capacity on National Forest Inventory and Monitoring systems (NFMS) needed to implement and better report on climate change including REDD+ and other improvements to forest data for management.

The Indonesian National Forest Inventory (NFI or IHN in Indonesian) has been implemented since 1989. FAO provided technical support with the World Bank funding from 1989 to 1995 to design and set up the system with MoEF and they were responsible for plot measurement after the initial project ended. The NFI data are essential for the MoEF for various needs: the MoEF regularly uses it to make a Country Report for the FAO's Global Forest Resource Assessment (FRA), and the data are also used data for deriving the national emission factors of various land covers when developing Indonesia's Forest Reference Emission Level (FREL) reports for UNFCCC (MoEF, 2015). The NFI cycle in Indonesia is 5 years by the law.

The original NFI aimed to estimate and report the volume, growth and development of stands according to forest type, species and groups of species in forest areas. It was not designed to provide estimates of forest carbon stocks but this information is now required by recent global reporting, including to UNFCCC. In addition, there is an increasing need for data and information to meet forest management and reporting requirements at national and global levels. The original design aimed for around 4,000 large 1 ha sample plots and became too expensive to measure and beyond the current resources and staff capacity to measure all of them every five years. So one of the objectives of this redesign was to reduce the costs and make a new design for the NFI to be more efficient and capable of being sustainably implemented by Indonesian government funds.

The Government Regulation Number 23, dated in 2021 and Regulation of the Minister of Environment and Forestry Number 7 (year 2021), mandate the forest inventory activities to include not only state forests, but also private forests and customary forests that were not included into the previous NFI design. Referring to these regulations and the growing needs for data and information, it is necessary to redesign the national forest inventory. In the context of these activities, there is a need to formulate the scope and objectives of the NFI to meet various requirements both at the national and global levels.

1.2 Purpose of NFI

The purpose of NFI is to record and assess the extent and quality of forests at the national level, in a timely, accurate and reproducible manner. Reliable, current and consistent information is required to inform domestic forest policy makers, implement national forest programmes, promote sustainable forest management, support forest research and fulfill national and international reporting commitments.

The focus of the NFI is at the national level (Figure 1) and not to inform local level forest management and planning. However, NFI methods and tools can be applied at the lower-level forest assessment and monitoring if the sampling intensity is increased to match the local conditions and requirements. Furthermore, NFI can provide support to local level planning by providing solid estimation methods and models.

Information provided by NFI can be summarized as follows:

- Forest area and changes (for instance Area of land, by land use classes and forest categories by land cover classes; Area of forest land by forest types, Area of forest land by tree species dominance);
- Volume and biomass of growing stock, including Trees Outside of Forests;
- Carbon and carbon pool changes of the five carbon pools (as in IPCC, 2003), i.e. Above ground, Below ground, Deadwood, Litter, and Soil carbon;
- Growth by tree species or tree species groups (through periodical monitoring);
- Condition of forests (untouched/damaged/degraded; managed/unmanaged, etc.);
- Biodiversity.

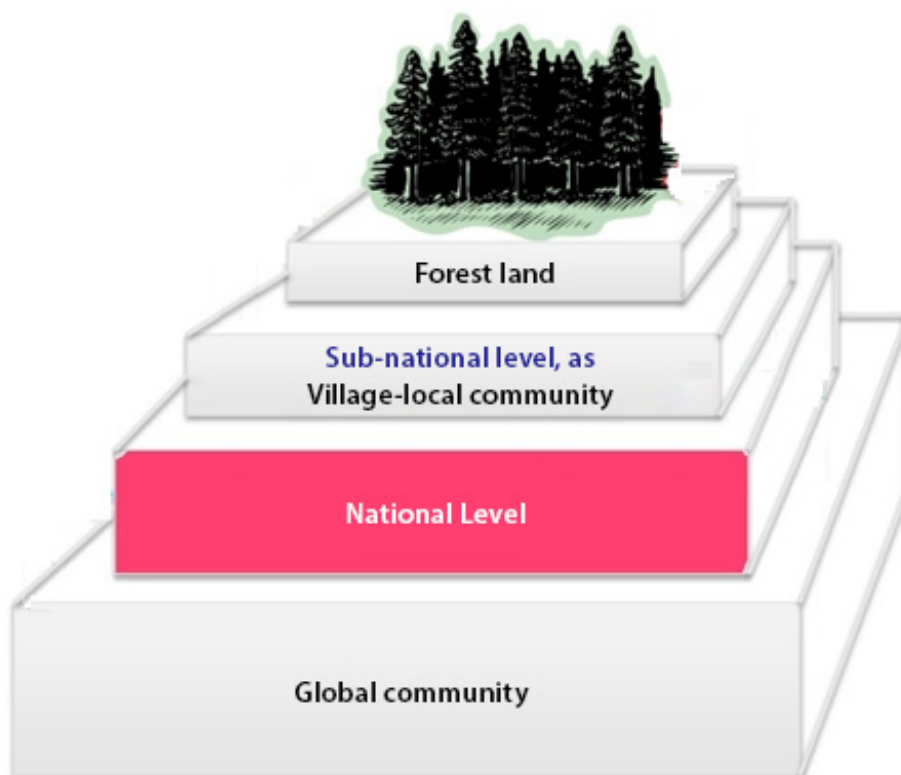


Figure 1. Focus of NFI is at the national level reporting.

NFI forms the base for a forest Monitoring, Reporting and Verification (MRV) system. One cycle of NFI will produce information with known accuracy on the area of land use classes and vegetation types, biomass of forests and biomass of trees outside forests. Repeated measurements will produce data on area changes by land use classes and changes in biomass, which can be used as a measure of degradation or gains. Detailed observations from the field combined with socio-economic data will facilitate analyses of the driving forces of forest degradation and provide information on utilization of forest products and services.

1.3 Basic concept

The objectives of an NFI are normally defined as a participatory process involving various end-users. The report of Tiriyana (2022) describes methods used in the previous NFI and current information needs of the MoEF.

Typically, relative importance of different attributes vary a lot and the list of needs to be fulfilled is rather difficult to limit while planning an NFI. Although NFI often has special and limited focus areas and “customers”, it can also provide results for indirect end-users. We present the following concept for making the decision how to formulate the final NFI plan:

1. **CORE NFI** contains the key variables (e.g. dbh, tree height, volume/ha, biomass, etc.) and field procedures that are consistent and uniform across all areas in Indonesia, with a target to cover information needs for islands (Appendix 3). This CORE is the framework for the NFI assessment; other inventory programs may be able to add (local) regional variables, but this should not change the CORE requirements.
2. **OPTIONAL NFI** covers those reporting areas and variables which are not required for getting the national level statistics, but can be recorded and reported by individual regions, provinces or other units; however, if specific variables (e.g. NTFP) are collected, they can be integrated into the same sampling scheme as ancillary variables. Typically, these are data to be collected for district level forest management planning or other purposes. Activities, such as a local Change Detection and Monitoring can be integrated to the same sampling system.

This concept allows defining different end-products and data collection schemes depending on the needs. It is also necessary to identify at which area levels the CORE NFI will provide accurate data and to which extent additional (optional NFI) data needs to be included into reporting. Optional NFI elements (such as smaller reporting areas, new attributes) can be defined and justified depending on actual needs and available resources.

Unless otherwise noted, all items in this document are considered to belong to the CORE NFI design. According to this design, the whole area of Indonesia is covered, including protected areas and other land uses. Optional NFI elements can be integrated to this framework but they will require additional resources including government and project staff plus more funding. The extra variables may possibly require extra time to measure the plots so we recommend caution to not expand the list and ensure a clear focus is put to complete the full NFI cycle in the planned 5 years.

When a NFI information system is established, the first task is the identification of information needs in terms of attributes, and area levels. The most critical factor is to decide the smallest area unit of interest and proceed with suitable forest inventory methodology. The minimum reporting area is used to define the extent and intensity of sampling. Considering that the main attributes are forest standing volume, biomass, and carbon, the target for the proposed design is to produce a sampling error less than 5 % for the mean biomass estimate for islands on drylands and 10% for mangrove stratum¹.

In Indonesia, the spatial forest structure is varying depending on the geographical locations and eco-physiological conditions. As a response to this, the results need to be calculated at the island level (Figure 2). Further details are outside of the scope of this strategic level planning.

¹ Additional analysis will be carried out in March 2023 to examine the consequences of setting in a 20 % accuracy limit (of the mean biomass estimate) for 6 main forest types by islands (see Chapter 2.2, excl. plantations).

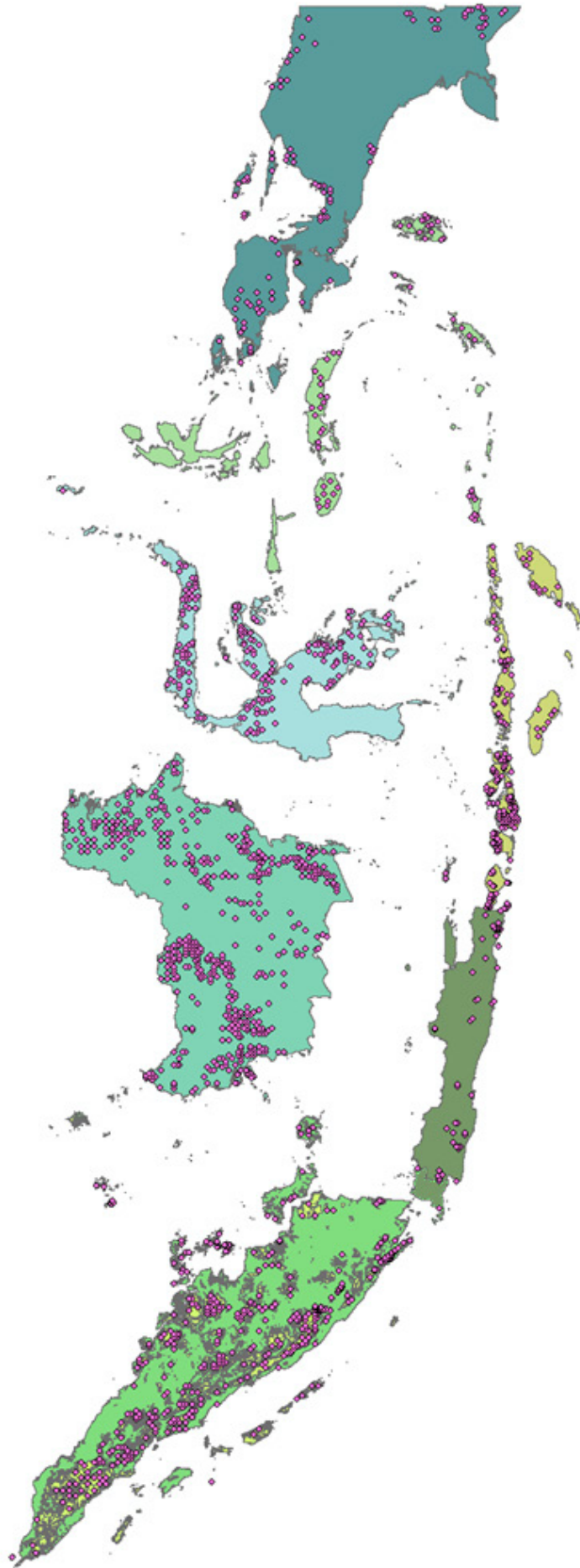


Figure 2. Islands and distribution of the subset of the past NFI field sample plots used in this NFI re-design analysis. The colours show the proposed 7 island groups for stratifying the new NFI.

1.4 Objectives

This report presents the initial results and findings of consultancies conducted by Senior Consultant Lauri Vesa and Prof. Timo Tokola, based on a 1-week mission to Jakarta and Bogor, and home-based consultancy conducted during January-February 2023. This assignment was carried out under the overall direction of Adam Gerrand (Chief Technical Advisor of the GCP/INS/087/NOR project) and Rubeta Andriani (National Project Manager), and in close collaboration with the FAOR Indonesia, Marco Piazza (Lead Technical Officer, LTO) and FAO's international and national Consultants. **The main purpose of this report is to examine and present alternative solutions for stratification and sampling design for the new NFI in Indonesia.**

It is assumed that islands of Indonesia are the final reporting units and calculation areas for forest inventory results (i.e., Sumatra, Jawa, Bali Nusa, Kalimantan, Sulawesi, Maluku, Papua).

The analytic works consisted of the following phases:

- A. Analysis of stratification alternatives:
Actions and goals: Geographic Information System (GIS) data review, and comparison of alternative stratification options and remote sensing (RS) data sources. The results should indicate a cost-efficient and up-to-date basis for the sampling frame. Results are documented in "Appendix 1: Review of alternative data and methods"
- B. Plot size evaluation:
Actions and goals: comparison of number of measured trees and percentages by diameter classes using a subset of the NFI data. The results should indicate if the efficiency of the plot system can be improved by changing the dimensions of nested plots, i.e. reducing measurements of small trees and increasing the proportion of large trees in samples. Results are documented in "Appendix 2: Analysis of Forest Structure and NFI Design"
- C. Sampling distance analysis:
Actions and goals: Analyze landscape structure with spatial similarity analyses and simulations with alternative NFI designs. The semivariogram and covariance functions quantify the assumption that things nearby tend to be more similar than things that are farther apart. The analysis results should indicate efficient distance between plots and cluster design. Results are documented in "Appendix 2: Analysis of Forest Structure and NFI Design"
- D. Cluster and plot accessibility analysis:
Actions and goals: Carry out tests with feasible sampling grids and assess realistic implementation strategies and outputs. The aim is to study whether GIS data and/or visual assessment of sampling units could produce reliable information on time consumption of field measurements so that this information could be used as a stratification criterion. Results are documented in "Appendix 2: Analysis of Forest Structure and NFI Design"
- E. Based on detailed analysis propose NFI design
Actions and goals: Present a feasible NFI design for Indonesia in the main part of this document.

2 Proposed Forest Inventory Design

2.1 Sampling designs options

The classical remote sensing related forest inventory designs are based on stratified, multi-phase or multistage concepts. Typically, cluster sampling is used in forest inventory and data collection when spatial correlation of forest landscape is low. When remote sensing -based alternatives are considered, main alternatives are multi-stage and multi-phase applications. Two-stage sampling is attractive when access to individual sample locations is time-consuming and costly. In two-stage sampling remote sensing data used only to stratify population into field tracts and RS data is not used to reduce variance like in multiphase designs.

Two phase sampling is typically used in forest inventories where remote sensing data covering the entire target area and field observations are combined. Large numbers of low-cost observations taken from quick visual interpretation of RS images (*i.e.*, Collect Earth) are used as auxiliary variables. Strata sizes are estimated by large first phase samples. In remote sensing, forest, and non-forest stratification don't need to be perfect, because estimation is based on actual field observations. Thus, the second phase observations always calibrate the upper-level strata size estimates. Two phase sampling can be extended to multiphase sampling if additional levels are planned.

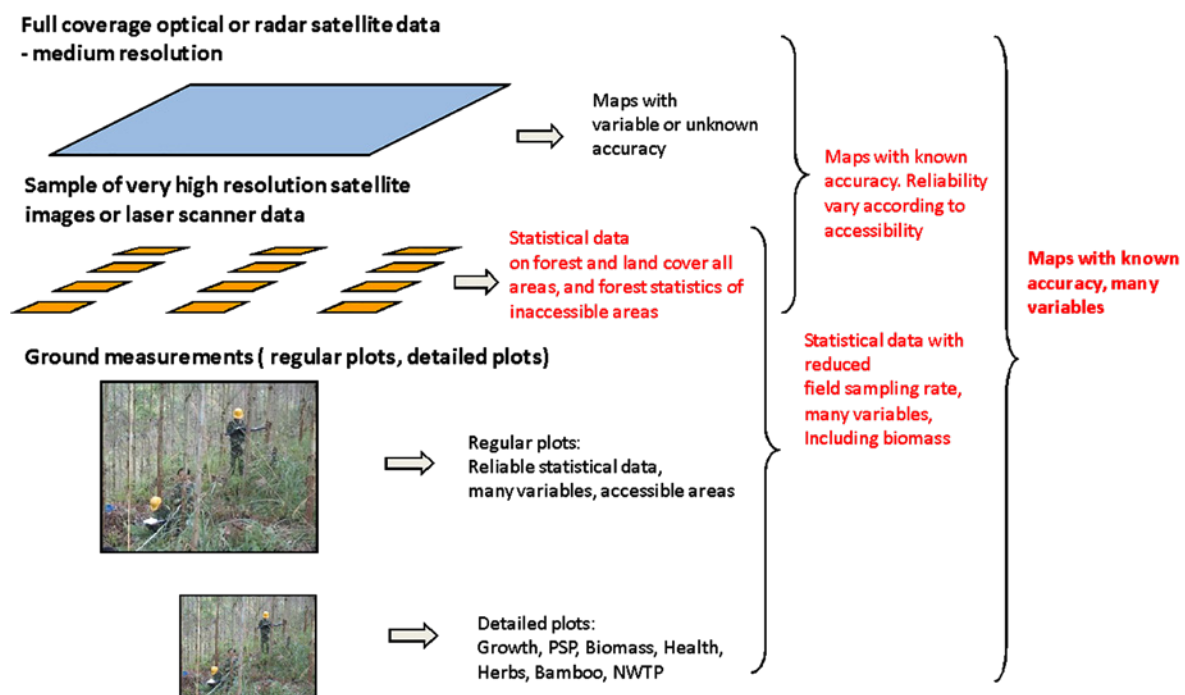


Figure 3. Multiphase sampling can be used to estimate forest attributes with remote sensing data.

In inventory and monitoring of forest resources at large scale, two-phase and three-phase designs are commonly used owing to their operational advantages (see, e.g., Schreuder *et al.* 1993). Usually, these multi-phase designs are accomplished on the basis of satellite imagery or aerial photo-based information collected on an intensive first phase sampling and subsequent phases performed by ground inspections in order to combine aerial and field data.

In a remote sensing -based forest inventory, at least two phases of sampling should be used. The first phase is based on full coverage of satellite image material and the main purpose is to stratify areas for more detailed observations. The second phase is based on Very-high Resolution (VHR) or Airborne Laser Scanning (ALS) data and phase three consists of measurements of growing stock (Figure 3).

2.2 Stratification and spatial arrangement of plots

Currently, there are three major forest habitats used by the MoEF classification in Indonesia: dry land forests, mangrove forests and swamp forests. Object delineation is based on visual interpretation of satellite data and done by digitizing on screen at a scale of 1:50,000. These are now further divided to primary and secondary forest areas:

1. Primary dryland forest,
2. Secondary dryland forest,
3. Primary mangrove forest,
4. Secondary mangrove forest,
5. Primary swamp forest,
6. Secondary swamp forest,
7. Plantation forest.

Besides the three habitat-based classes, the NFI design approach should also consider three strata which will have consequences for the sampling design as well as for the logistics of the field implementation: trees outside forests (TOF), uncertain forest and “inaccessible” forest areas. Uncertain areas are categorized when we assume that a field control is required for assessing the true stratum value. Failure to acknowledge uncertainty in forest/non-forest classification can result in biased and misleading inference on the forest variable of interest (see Finley *et al.*, 2011; Magnussen *et al.*, 2012). Inaccessible areas require alternative remote sensing based (high resolution) assessments. The inaccessibility is defined by the following factors: walking distance (i.e. more than 5 km to a road or navigable river), steep terrain or unsafety of field work (e.g. political instability).

The current forest area definition of Indonesia is rather strict (i.e. “a land area of more than 6.25 ha with trees higher than 5 meters at maturity and a canopy cover of more than 30 percent”) and thus leaving many smaller forest patches out from the official statistics. Current satellite image materials (e.g. Landsat, Sentinel) also allow interpretation of 1 ha areas. Government should consider changing the area criteria from 6.25 ha to 1 ha. However, in this document we still used current definitions, because there is no indication for change. Hence a separate “trees outside the forest” assessment to fully assess the extent of trees and tree biomass also in areas not classified as forests can be carried out and it can be based on a sample grid to assess the amount of scattered forest patches. If intensified image interpretations could be done outside of forest areas, inventory of “trees outside the forest” can be integrated to interpretation work done at the first phase sampling of the NFI.

When very detailed ecological categories are defined to form strata, the number of required field plots quickly gets to a very large number to produce statistically reliable estimates. This has a big effect on increasing the time and expense of plot measurement. It also complicates the analysis, especially of change data where the classes get multiplied and large numbers of classes with small plot numbers mean the results have higher uncertainty or in some cases not enough plots to be able to determine if there has been significant changes over time.

The vegetation cover types vary across Islands, so number of plots by strata and design of first phase sample is defined by Islands. Secondly, a uniform sampling design for all forest types would mean too few plots in the Mangrove stratum, which is known to contain large carbon stocks particularly in soil. Thus, we recommend using two main forest classes (1 and 2) and four other strata to form the sampling strata for each Island, as follows:

1. Dryland and swamp forests (canopy cover >30%, area > 1 ha tbc, currently 6.25 ha),
2. Mangrove forest (canopy cover >30%, area > 1 ha tbc, currently 6.25 ha),
3. Uncertain forest (canopy cover >30%, area > 1 ha tbc, currently 6.25 ha), these are areas that cannot be clearly identified as being forest or not in the pre-field checks using remote sensing.
4. Trees outside forests (canopy cover 10–30%, area from 1.0 ha to 6.25 ha),
5. Inaccessible areas (separate ground truthing with detailed remote sensing analysis),
6. Other land uses.

The reason for separating two forest strata from other four categories is that these are expected to be quite different in terms of vegetation cover and amount of biomass.

During the first phase of the forest inventory, the proposed approach needs to interpret auxiliary GIS data, define stratum value for each grid point and locate forest plots to be measured in the field. This is done for these first phase plots defined in the basic sample grid. Stratum values are either visually estimated with auxiliary material, or suitable GIS layers are used directly and attached to grid points (*i.e.* extract values to the grid points).

The following auxiliary datasets can be utilized:

- Land cover and Land use maps, Mangrove survey & Peatland surveys or maps;
- Tree height and tree cover maps based on high resolution data (e.g. Lang *et al.*, 2022);
- Accessibility maps, based on Digital Elevation Model (DEM), Distance to roads, Distance to navigable rivers (described with polygons in national GIS maps);
- Global Ecosystem Dynamics Investigation (GEDI; see Dubayah *et al.* 2022, Zhang *et al.* 2022) global canopy height data, can be used to stratify remote areas;
- Airborne Lidar/laser data (ALS) data can be used as a proxy to replace field observations.

Tree height maps indicate directly the canopy height of potential vegetation to be assessed. Land cover map is done to describe land use and combined with extra map materials. Peatlands, Mangroves and many non-forest areas are mapped with detailed remote sensing data and field visits. It is worthwhile to use this information in the assessment of “forest/non-forest/uncertain” -classification. Figure 4 demonstrates a hexagonal grid overlaid with land use map and satellite image-based forest area (dark green).

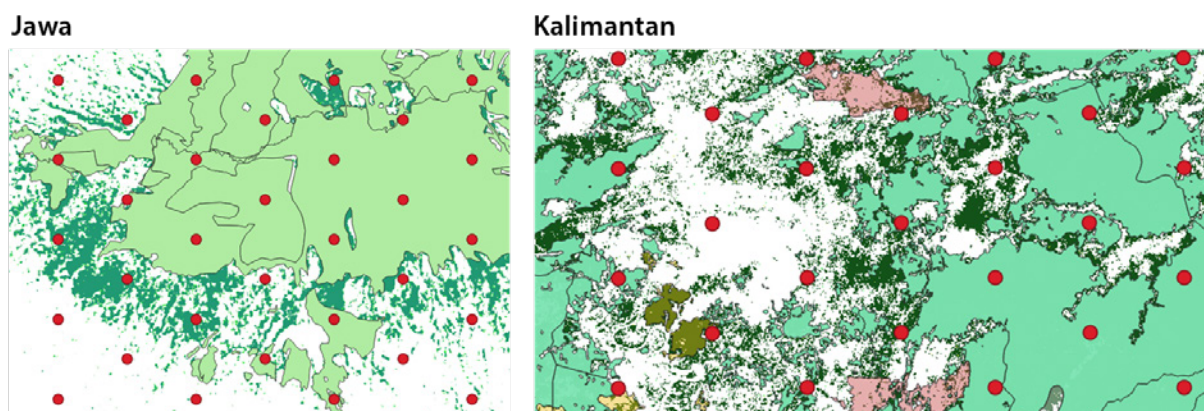


Figure 4. Examples of hexagonal grids overlaid with land use map and satellite image-based forest area (dark green) in Jawa (5 km grid) and Kalimantan (20 km grid).

The historical criterion for minimum forest area is 6.25 ha in Indonesia although 1.0 ha is becoming increasingly common in international reporting. So depending on the Minimum Mapping Unit threshold decided for the new monitoring system then GIS commands can be used to remove too small patches out from raster maps (e.g., down to 10 m resolution pixel size for Sentinel).

2.3 Plot design

Introducing nested circular plots (instead of a square 1-ha plot as in the the previous NFI design for Indonesia) is an approach that is typically employed in large scale forest inventory because it reduces the total number of trees to be inventoried (thus minimizing time and cost of implementation) without relevant loss in the accuracy of volume, biomass or carbon stock estimates. More specifically, the reasons for this are listed as follows:

- Circular plots are efficient because the actual boundary around the plot does not need to be marked, and only one ground reference point (center of the circular plot) is enough.
- If permanent plots are used, it is necessary to identify each measured tree for remeasurement. With circular plots the location of any tree is easy to assess and record by measuring the distance and bearing from the plot center. If this is done well then it may not be necessary to tag or number in the field which saves time and reduces the attention to the plot and reduces disturbance.
- Especially, when electronic measuring tools are used for measuring distances, it is fast to measure when a tree is “in” or “out”².
- Circular shape minimizes the occurrence of “border-trees” as compared to a rectangular polygon of the same size.

A nested plot design is widely applied in forest inventories globally (e.g. FIA in the USA, NFI and FMU forest inventories in Indonesia). In general, a nested plot reduces the sampling effort for smaller trees in favor of larger trees that contribute a greater share to total volume and biomass in forest (e.g., Kershaw *et al.*, 2016). This is why fixed-area plots (without nesting) may be less efficient in stands with a large number of small trees, because of the additional time required to measure them and it is more efficient to measure tree data evenly across diameter distribution.

The proposed plot designs for different variables are described in Table 1 and 2. They are based on intensive analysis of the sub-set of 1,400 sample plot data that FAO had access to and some of the analysis shown in Appendix 2.

All field plots are considered permanent, and these are marked with a metal peg/post often driven into the ground to be hard to remove as the plot reference point (i.e. usually plot's center point) coordinates are recorded by averaging several minutes of Global Positioning System (GPS) measurements. The reason for establishing permanent plots is that they are far more efficient in estimating changes (land use changes, changes in biomass) than temporary plots.

In case the sub-plot area covers more than one land use/vegetation (LUV) type, the plot can be divided into sections representing homogenous land use or vegetation type units, with variable size and shape that have been identified in the field (Figure 5). The share is presented as percent of the total plot area for each nested plot.

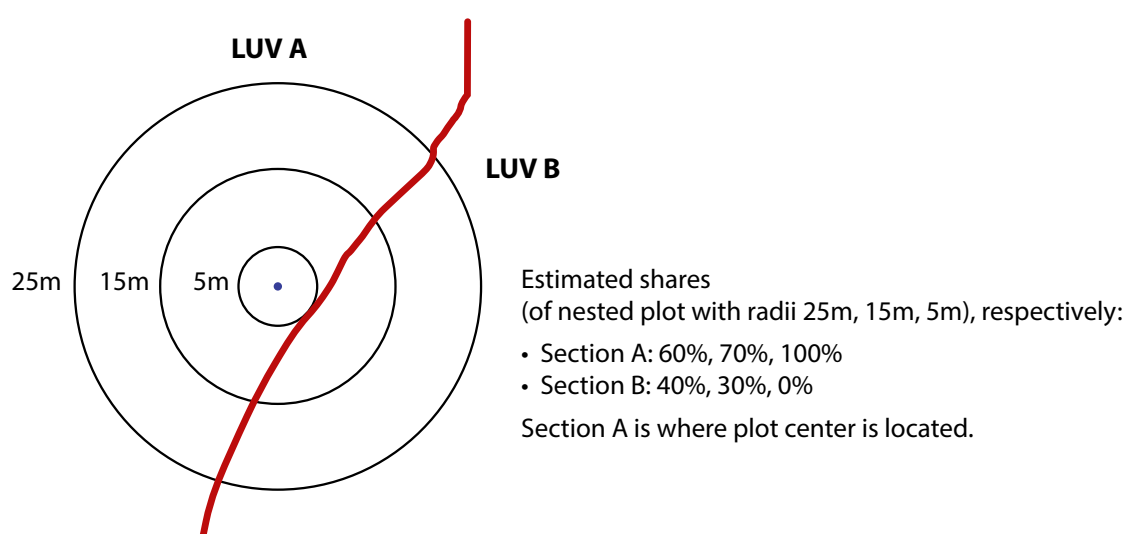
² Electronic measuring tools like Haglöf Vertex (ultrasound) and TruPulse Rangefinder (laser) have built-in slope correction, so that correct distance is easy to define. Both Vertex and TruPulse Rangefinder can be used for distance and tree height measurements.

Table 1. Survey unit specification, circular plot design on drylands.

Unit name	Radius/Dimensions	Measurements
Plot	25 m	Trees: DBH > 40 cm
Subplot 1	15 m	Trees: DBH > 20 cm Bamboo: DBH > 10 cm Dead wood (lying): diameter > 10 cm Stumps: diameter > 10 cm
Subplot 2	5 m	Trees: DBH > 5 cm Bamboo: DBH > 2 cm Dead wood (lying): diameter > 2 cm Climbers: diameter > 2 cm
Optional: Subplot 3	2 x 1 m x 1 m	Saplings: 1 cm < DBH < 5 cm Seedlings
Optional: Soils sampling	soil pit	
Optional: Litter and Coarse Woody Debris	2 x 1 m x 1 m	

Table 2. Survey unit specification, circular plot design for Mangrove stratum.

Unit name	Radius/Dimensions	Measurements
Plot	12 m	Trees: DBH > 20 cm
Subplot 1	8 m	Trees: DBH > 10 cm Dead wood (lying): diameter > 10 cm Stumps: diameter > 10 cm
Subplot 2	5 m	Trees: DBH > 5 cm Dead wood (lying): diameter > 2 cm
Subplot 3	2 x 1 m x 1 m	Saplings: 1 cm < DBH < 5 cm Seedlings
Optional: Litter and soil composite samples	In Field manual for soil and litter sampling	

**Figure 5. Example of land use/forest type sections distribution within a nested circular plot.**

2.4 Cluster design

Cluster sampling is widely used in large area forest inventories, such as in NFIs. The major reason does not lie in the statistical performance but in practical aspects: once a field team has reached a field plot location (which in some cases may take many hours or even some days), then one wishes to collect as much information as possible at this particular sample point. The major point in cluster sampling is that for each random selection not only one single sampling element is selected but a set (cluster) of sampling elements; thus, a cluster consists of a group of observation units, which together form a sampling unit. The spatial configuration of cluster plots (in terms of number of subplots and their distance) is a matter of optimization based on the Intraclass Correlation Coefficient³ (ICC) and practical considerations on costs. The concept for calculating cluster sampling-based sample size is based on ICC as following:

- Step 1: Calculate Simple Random Sample Size (*SRS*),
- Step 2: Calculate ICC – Calculated *rl*, Number of Clusters,
- Step 3: Calculate Design Effect, $DE=1+(n-1)rl$,
where *n* = number of plots in cluster,
rl = intraclass correlation coefficient,
- Step 4: Multiply *SRS* with *DE*,
- Step 5: Calculate the minimum number of clusters required.

This means that cluster sampling carries a larger error variance than simple random sampling with the same number of subplots. The planner, however, strives to keep the intraclass correlation as small as possible in order not to lose too much precision. A spatial sampling simulator⁴ was used to generate alternative cluster designs over GIS raster layer (i.e. ETH canopy height estimates, see details in Appendix 1 and example in Figure 6 with two different cluster options).

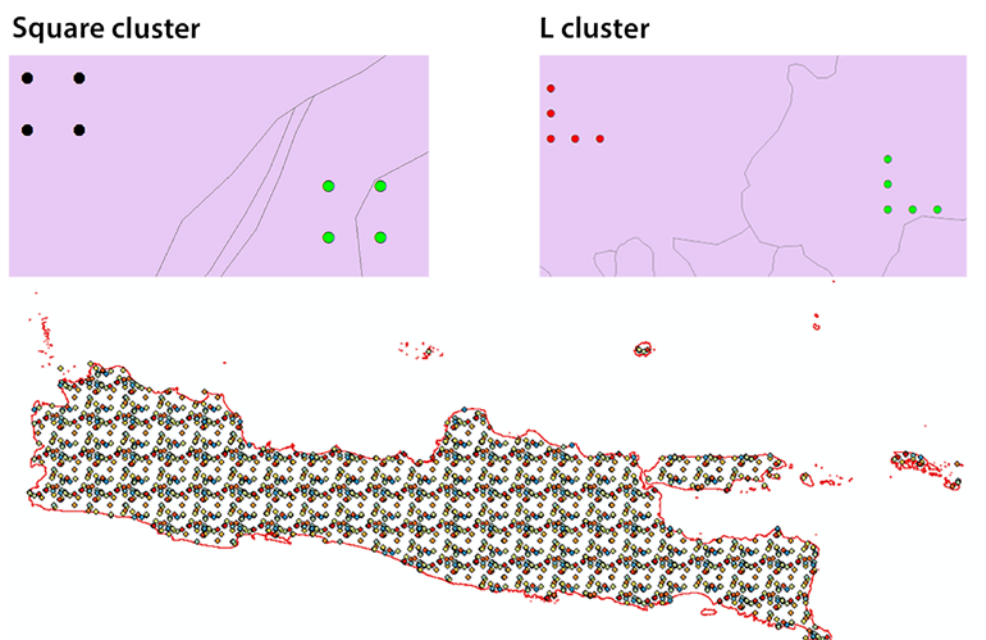


Figure 6. Example of regular pattern grid generated to study intraclass correlation and define feasible spatial design.

³ See e.g. http://wiki.awf.forst.uni-goettingen.de/wiki/index.php/Intraclass_Correlation_Coefficient

⁴ <https://openforis-shiny.shinyapps.io/plot-generator/>

In general, it is not possible to identify one cluster shape as generally superior, as the cluster shape also relies on the forest inventory cost to ensure that it will be practical in the field. From a purely statistical point of view, L-shapes are good because of their larger spatial extension and the larger average distances between pairs of subplots which will lead to a lower intra-cluster correlation on average (Kleinn, 1996).

Considering that the preferred approach is a cluster of samples, alternative approaches were tested with simulation study (see Appendix 1). Various cluster sizes and shapes were tested (1–5 samples in a cluster). The next major parameter to solve was the appropriate distance between plots. When intracluster correlations were calculated for different design options for Islands, it was clear that short distances are producing a lower number of samples. It was concluded that IICs were high and therefore clustering of plots is not effective in reasonable distances (in terms of practical field work). It means that single large plots or few subplots with short distances is the most efficient design.

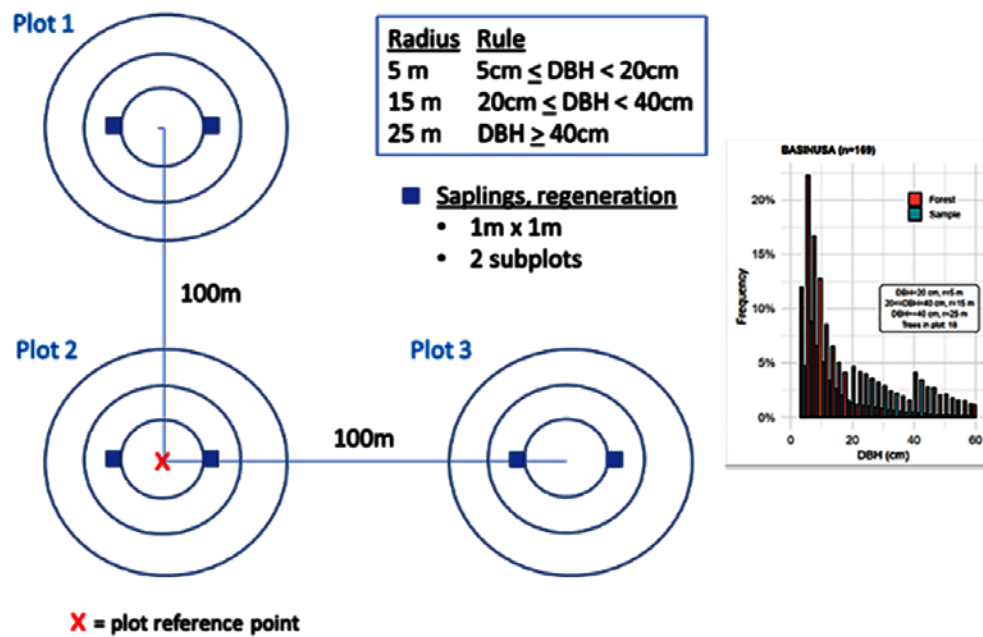
When distances of 250–350 meters between samples are applied, the total number of samples would be 9,529. When a shorter distance (from 50 m to 100 m) is applied, the total number of samples is 8,481 (Table 3)⁵. So the shorter distances can provide a more effective design, and applying larger sample plot sizes can be justified (Figure 7 and 8).

Table 3. Comparison of small plot design (plots in tracks without ICC) and larger plot design (3 subplots, correlated) in different Islands.

Concept		N. Plots required			Plotarea	
ISLAND	Legend	10%	5%	DE*n	without DE	with DE
BaliNusa	r = 25m (1 subplot)	156	614	1109	614	1109
BaliNusa	r = 25m (3 subplot)	119	476		1428	1428
Jawa	r = 25m (1 subplot)	209	836	1583	836	1583
Jawa	r = 25m (3 subplot)	148	590		1770	1770
Kalimantan	r = 25m (1 subplot)	135	537	1452	537	1452
Kalimantan	r = 25m (3 subplot)	105	413		1239	1239
Maluku	r = 25m (1 subplot)	138	552	1289	552	1289
Maluku	r = 25m (3 subplot)	55	219		657	657
Papua	r = 25m (1 subplot)	100	398	1024	398	1024
Papua	r = 25m (3 subplot)	65	257		771	771
Sulawesi	r = 25m (1 subplot)	165	658	1807	658	1807
Sulawesi	r = 25m (3 subplot)	121	486		1458	1458
Sumatera	r = 25m (1 subplot)	132	525	1265	525	1265
Sumatera	r = 25m (3 subplot)	97	386		1158	1158
					Large Plot	8481
					Small	4120
						9529

⁵ These numbers of plots do not refer to the proposed number of samples in the design, but these are results derived from this analysis.

TREES: Drylands



TREES: Mangroves

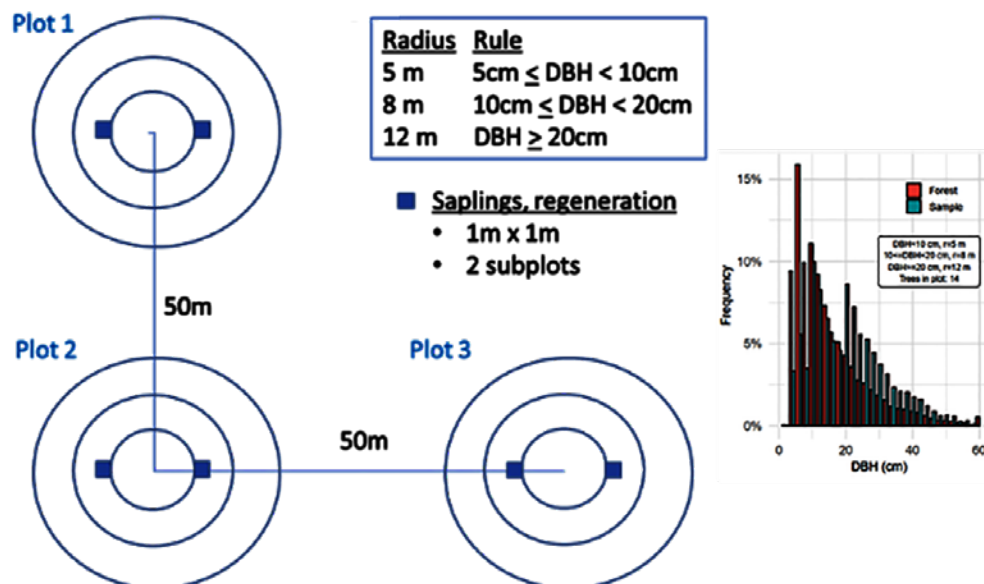
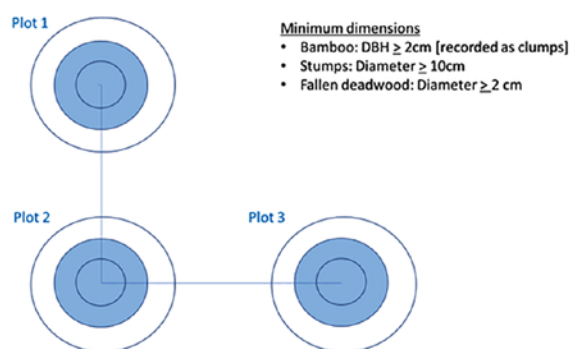
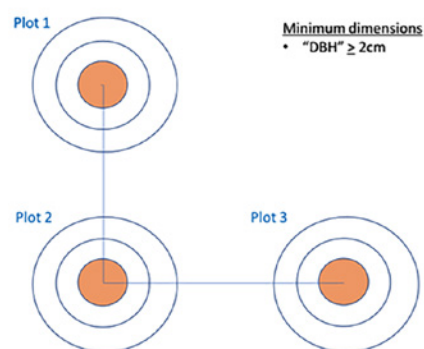


Figure 7. Layout of proposed plot design (trees).

BAMBOOS, STUMPS, FALLEN DEADWOOD



CLIMBERS/RATTAN



OPTIONAL: LITTER, COARSE WOODY DEBRIS

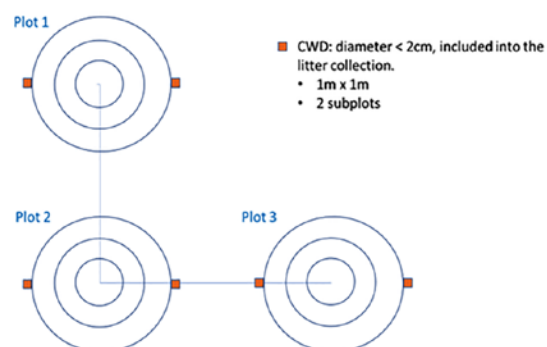


Figure 8. Bamboo, stumps, deadwood, rattan and optional objects to be recorded in plots.

The required number samples (i.e. a plot consisting of 3 subplots) was calculated with the help of old NFI field plot data using the Cochran's formula with given accuracies (Table 4). The number of samples per island is defined by using variations calculated from the subset of NFI Permanent Sample Plot (PSP) data (see Appendix 2). A total of 1,400 NFI PSPs from the total of 4,000 PSPs were used in the analysis.

We expect that the proposed design produces a sampling error less than 2 % for the forest area estimate and around 2 % for the mean biomass estimate for the whole country. The target sampling error on drylands strata is 5%, and 10% for mangrove forests for the mean biomass estimate.

Table 4. Estimated number of samples by islands with preferred cluster plot designs. Sample size for Mangroves is computed for the whole country.

ISLAND	Plot design	Accuracy	n
BaliNusa	r = 25m (3 subplot)	5%	476
Jawa	r = 25m (3 subplot)	5%	590
Kalimantan	r = 25m (3 subplot)	5%	413
Maluku	r = 25m (3 subplot)	5%	219
Papua	r = 25m (3 subplot)	5%	257
Sulawesi	r = 25m (3 subplot)	5%	486
Sumatera	r = 25m (3 subplot)	5%	386
Mangrove	r = 12m (3 subplot)	10%	192
Total			3019

The proposed sampling unit design (a plot with 3 subplots) was compared with 1-ha plot (see estimated number of samples and cumulative sampling area to access the desired accuracies in Table 5). The savings in terms of sampling area is about 34% compared to the 1-ha PSPs.

Table 5. The proposed sampling unit design compared with 1-ha PSPs.

	Sampling error			
	20%	15%	10%	5%
Number of samples	178	318	710	2827
Number of samples compared to 1 ha plots	113%	117%	117%	117%
Sampling area compared to 1ha plots	64%	66%	66%	66%

2.5 Sampling frame

The number of samples and spacing was determined according to the forest area of each island. Table 6 shows how sampling unit spacing was calculated. In this calculation, a predefined number of sampling units is allocated to forest areas inside each island. The spacing is then calculated in a double density of samples to ensure that these are scattered enough, but so that there is still the possibility to find enough accessible sampling units. This will ensure that there is flexibility, when uncertain and inaccessible areas are delineated within forested areas.

Papua is proposed to have a Global hexagon with spacing of 29 km. Kalimantan, and Sumatra, are proposed to have a 17 km Discrete Global Grid (DGGRID), while Sulawesi and Maluku are proposed to have a 9.8 km spacing grid (resolutions 11 and 12). The densest spacing (5.6 km) is proposed to Jawa and Bali Nusa (Table 6).

Both regular and hexagonal types of grids were generated and laid over the area of each island. These grids are needed to interpret stratum values using auxiliary GIS and RS information. Figure 9 shows the difference between a regular and hexagonal grid in Sumatra. The actual difference is very small and we propose using hexagonal grids with varying distances by islands.

A fixed number of plots can be applied for specific strata (mangrove, trees outside of forests). In mangrove sites, maps of mangrove areas can be applied with a 200 m buffer zone to capture thin coastal slivers of land caused by digitizing errors. We recommend using a two-phase sampling design: original grid intensity (of the island) will be intensified to 1 km x 1 km grid for visual interpretation, and in the second phase altogether 30 field plots are allocated randomly to each island.

Similarly trees outside forest areas (i.e., canopy cover 10–30%, area 1–6.25 ha), we propose to establish 30 field plots to each island.

All the plots belong to the first phase sample but those plots that according to prior information (recent very high-resolution remote sensing data) without doubt are in the forest are directly assigned to be second phase plots. Plots falling outside forest areas are assigned to belong to the non-forest stratum. If a certain location contains trees or forest status is somehow uncertain, that should be assessed in the field.

Table 6. Proposed spacing when Land Use forest maps are used in stratification. Number of plots in proposed hexagon grids vs original plan (N=Number of Hexagon Plots, t-km2=1000 km2 in Land Use Map) and distribution of plots in Land Use categories.

Landuse/Landcover	Forest/Non-Forest	BaliNusa		Jawa		Kalimantan		Maluku		Papua		Sulawesi		Sumatera		Total	Total
		N	t-km2	N	t-km2	N	t-km2	N	t-km2	N	t-km2	N	t-km2	N	t-km2	N	t-km2
Primary dryland forest	forest	403	8.7	24	0.6	305	94.7	72	6.0	438	191.6	559	46.4	574	49.4	2375	397.4
Secondary dryland forest	forest	914	17.7	375	8.0	404	131.0	674	40.6	184	74.1	589	43.6	665	45.3	3805	360.2
Primary mangrove forest	forest	11	0.1	6	0.1	2	1.6	46	2.6	36	12.9	23	0.3	39	1.5	163	19.3
Primary swamp forest	forest					1	0.6	2		84	44.7	1	0.1	27	2.3	115	47.7
Plantation forest	forest	14	0.2	839	18.7	35	10.1	6	0.4		0.1	4	0.4	278	22.8	1176	52.7
Secondary mangrove forest	forest	27	0.2	16	0.3	19	4.6	39	1.5	5	2.4	52	1.1	109	4.4	267	14.6
Secondary swam forest	forest			1	0.0	103	32.1	5	0.2	50	18.1	1	0.1	158	12.5	318	63.1
Dry shrub	nonforest	404	6.1	53	1.2	111	37.1	229	9.2	26	9.0	288	16.0	393	24.5	1504	103.1
Estate crop	nonforest	12	0.2	210	4.5	224	70.2	5	0.4	10	3.0	102	6.4	1551	125.2	2114	209.9
Settlement areas	nonforest	174	2.9	738	16.5	14	3.4	10	0.5	2	1.4	62	3.1	173	13.0	1173	40.8
Bare ground	nonforest	43	0.3	14	0.3	18	6.4	7	0.4	8	3.6	13	0.4	51	3.0	154	14.3
Savanna and grasses	nonforest	680	11.4	11	0.3			35	1.4	27	10.3	84	3.6	13	1.0	850	28.0
Open water	nonforest	27	0.4	29	0.8	17	4.1	11	0.2	16	6.4	53	2.3	79	5.4	232	19.5
Wet shrub	nonforest	3	0.0		0.0	118	35.8	6	0.2	41	18.2	4	0.4	297	21.5	469	76.1
Pure agriculture	nonforest	478	7.5	762	17.1	27	6.9	71	2.3	3	0.7	223	13.9	405	33.6	1969	82.1
Mixed dry agriculture	nonforest	598	10.7	1253	27.8	234	74.4	322	11.6	18	9.8	567	32.7	1126	87.1	4118	254.2
Paddy field	nonforest	362	6.7	1677	36.2	29	7.4	5	0.3	1	0.6	188	13.1	230	16.8	2492	81.1
Fish pond/aquaculture	nonforest	21	0.3	97	2.2	15	3.3		0.0		0.0	47	2.2	44	1.9	224	9.9
Port and harbor	nonforest	2	0.0	4	0.1	1	0.0	1	0.0	1	0.0		0.0		0.1	9	0.3
Transmigration areas	nonforest	1	0.0			1	0.5	3	0.1	1	0.5	2	0.2	9	0.7	17	1.9
Mining areas	nonforest	2	0.0	5	0.1	14	5.1	3	0.1	10	0.0	5	0.4	44	2.7	83	8.5
Open swamp	nonforest	1	0.0156		0.0	28	6.1		0.0		7.3	1	0.0	34	2.1	64	15.5
Total, Forest, Hexagon (2 x original spacing)		1369		1261		869		844		797		1229		1850		8219	
Total, Forest, planned theoretically with mangrove		506		620		443		249		287		516		416		3037	
Total Mangrove		30		30		30		30		30		30		30		210	
Ratio, Hexagon/planned		2.6		1.9		1.8		3.0		2.5		2.3		4.1		2.5	

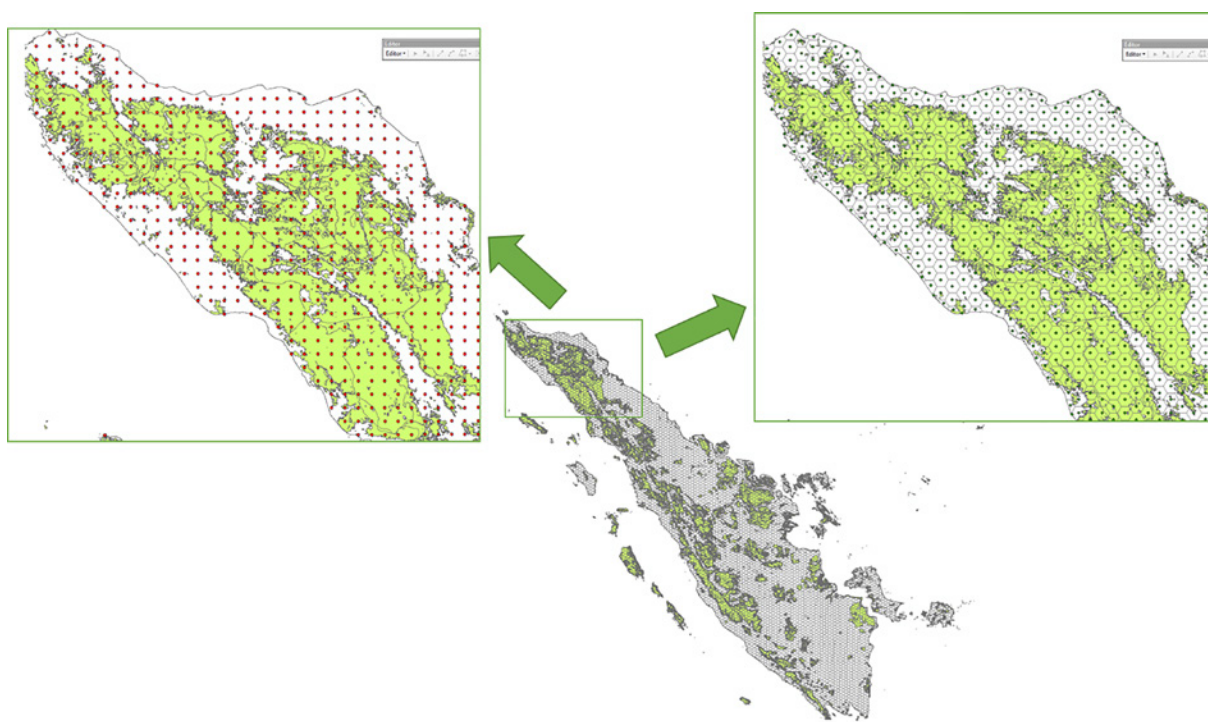


Figure 9. Illustration of regular grid and hexagonal grids in Sumatra.

When the plots were distributed into maps and overlaid with the existing Land use map, we can calculate distribution of plots in specific land use categories. When land use category was rare in certain Islands, it is natural that only few plots were falling to that category. In this inventory design, forest stratum contain all dryland areas, plantations and peatland areas. Mangrove areas were separated to have their own fixed number of plots (30). The planned GRIDS had oversampling (15 – 2.7 times depending on Island). The purpose of oversampling was to have some flexibility in terms of inaccessibility and to ensure that enough predefined plot-locations are available for the second phase field plots.

2.6 Accessibility of sampling sites

During the NFI cycle 1991–1999, it was estimated that the extent of accessible area was around 7.5 mill. ha and inaccessible areas cover a total of 16.2 mill. ha. A rapid GIS analysis was completed to investigate the current situation (with the help of an existing Digital Elevation Model and road maps). Two criteria were tested in the GIS analysis, as follows:

1. elevation is above 1000 m, and
2. distance to roads and major rivers is greater than 5–12 km.

An access difficulty factor can be calculated to each proposed plot (Figure 10 and Figure 11) and selection of samples for field measurements can utilize this layer. Proportions of inaccessible areas are significant particularly in Kalimantan and Papua where, in addition to difficulties caused by terrain and distances, the occasional local conflicts can affect accessibility. Moreover, due to security issues, there are also areas in Aceh and Central Sulawesi where it can be difficult to access the planned sampling sites. However, these assumptions will require further verifications.

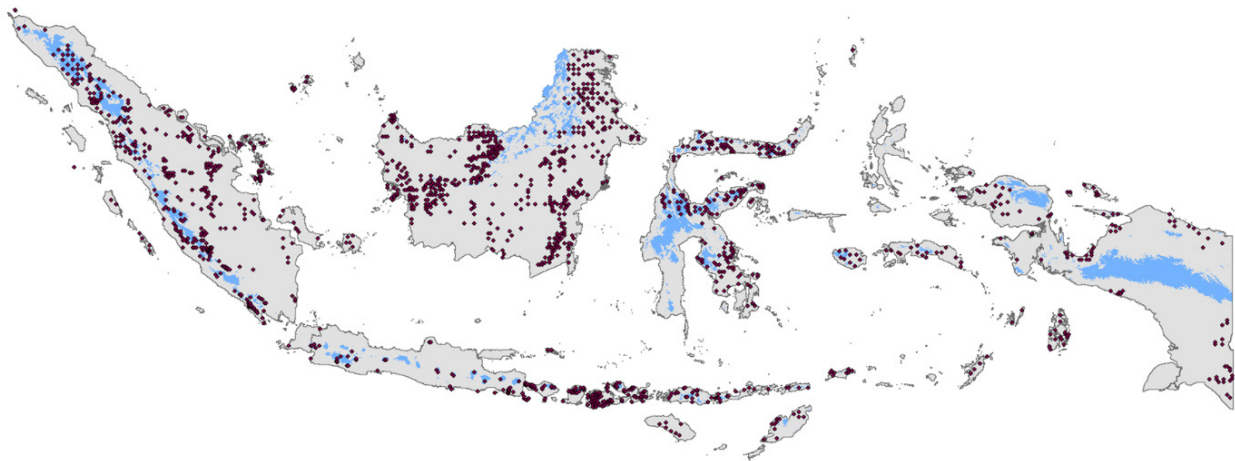


Figure 10. Accessibility analysis based on elevation.

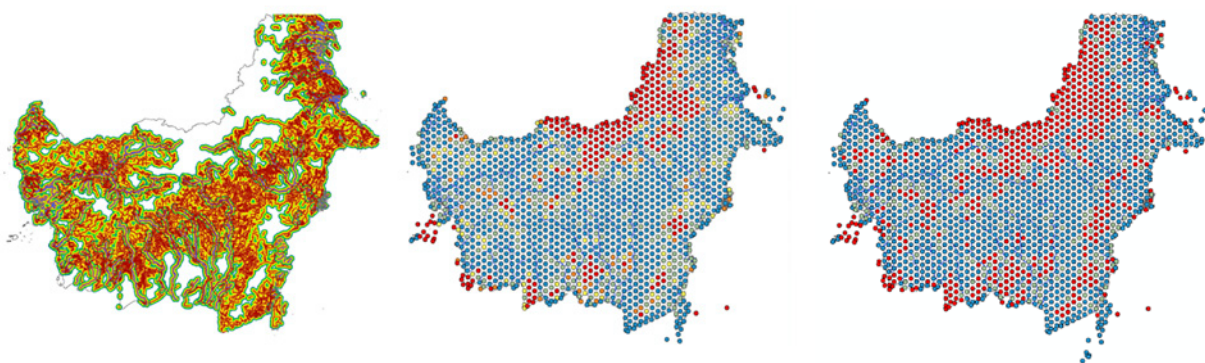


Figure 11. Distance to roads and rivers in Kalimantan (left). The shortest distance was computed for each cluster's reference point. Red color indicates over 12 km distances (middle) and 5 km distances (right).

The large inaccessible areas need to be covered by using an airborne LiDAR to replace field observations. GEDI observations could also be utilized, but reliability of airborne LiDAR would yield much better reliability than the spaceborne version. A few sample lines are enough to assess the amount of forest biomass in these areas (Figure 12).

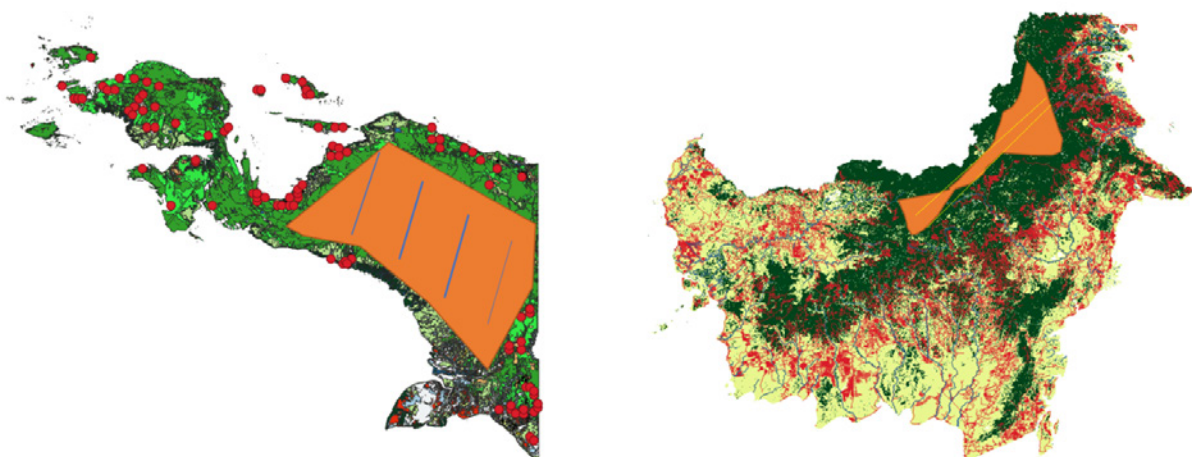


Figure 12. Areas to be covered with Airborne Laser Scanning (ALS) sample lines (5-10 lines per area).

2.7 Cost estimates

All field plots are permanent. The NFI variables are split into two categories: core variables and optional variables. The core variables will be collected following field procedures that are consistent and uniform across all areas in Indonesia. The optional variables will be collected either by a subset of sample sites or differently by strata. Optionally, soil and litter samples can be collected in a subset of plots, and samples are analyzed in a soil laboratory in Indonesia. Optional variables are not included into the following cost estimates.

We recommend collecting socio-economic data in a campaign that is separate from NFI because the effective sampling design for socio-economic data collection is different from the design that is planned for collecting biophysical data.

Normally, estimation of time consumption for the field assessment is the first step in calculating the total cost for an NFI. Since 1997, the NFI implementation has been based on national funding. According to Tatang (2022), the average unit cost of USD 3,666 was required to cover all expenses for enumerating one TSP/PSP plot, and the latest cost estimate was USD 3,960, covering team meetings, travel costs, meals, as well as honoraria. A survey team consists of BPKH staff (usually three people) and local workers (usually 2–4 people). Depending on cluster plot locations, a survey team needed 18–22 days to establish and measure one new TSP/PSP plot and 7–14 days to re-enumerate a former PSP. In the new design, the plot area (with 3 sub-plots) is approximately 57% of NFI PSP's area and fewer small trees are measured from that area. It can be assumed that one sampling unit (i.e., a plot with 3 sub-plots) can be measured within 1–2 days.

The estimated costs for an entire forest inventory campaign covering the whole country are presented in Table 7.

Table 7. Estimated total NFI costs.

Item	N	Cost per Plot, USD	Total, USD
Field Plots	3037	2250	6 833 250.00
Control Plots			1000 000.00
TOF plots	210	1000	
Lidar	1		1000 000.00
RS&GIS work	1		250 000.00
Administration	1		1300 000.00
Total			10 593 250.00

The estimated cost of field assessment is USD 7 million, and with possible Quality Control teams the total cost is USD 8 million. In addition, the NFI project will require more investments (summarized in Table 8). Based on these estimates, the total cost of a full scale NFI is in the range USD 11–12 million. These cost estimates do not include development of allometric models, nor other research activities. The cost and quality of remote sensing data are normally connected to each other. Feasible satellite data (such as Sentinel or Landsat) are free and we can easily cover the entire country with low material and interpretation cost. GEDI data is also free and it can be used in biomass predictions and small area estimation. ALS data is costly but can be utilized on a sample basis in inaccessible areas.

Table 8. Estimated additional cost of a NFI project

Cost item	Estimated cost, USD
Field equipment	100,000
IT equipment, tablets	60,000
RS materials (i.e. satellite images)	40,000
Other materials (e.g. printed maps)	30,000
1 st phase sampling (mangrove)	10,000
Training	100,000
Data entry, data management	30,000
Ancillary costs	70,000
Administration	15% (~876,000)
Total	1,316,000

2.8 Biodiversity, trees outside of forest and non-timber forest products

Assessment of biodiversity and non-timber forest products (NTFPs) are normally integrated as a part of NFI. 'Trees outside of forest' (TOF) mapping is mainly done outside forest areas. There is a clear need for further integration of these both issues to regular NFI. Biodiversity and wood consumption statistics (as a part of socio-economic studies) needs further integration and additional analysis using habitat mapping and external data sets. TOF area mapping could be integrated also into the same NFI image interpretation process, because in the NFI we need to define forest areas, and at the same time we can also interpret TOF areas.

Data collection of biodiversity information will be made partly through sampling approach and partly through public participation, with help of governmental guidance. Many countries have started to utilize the international facility called "Global Biodiversity Information Facility" (GBIF), which is developing a set of core services to support global integration of biodiversity data, based on the GBIF Biodiversity Data Architecture.

Vegetation composition of the country is derived from multiple factors including geography and climate. Both horizontal (north to south) and vertical (tropical to temperate) variations contribute to the diversity and uniqueness of ecosystem, flora and fauna along with different climates. The zoning of these ecological areas is important to give a broader idea of the vegetation distributions in the areas. These aspects should be taken into account for the biodiversity assessment process.

Distribution of the various NTFPs follows the physiographic condition of the areas. Ecological mapping will give clear ideas about the distribution of the species.

Prioritization of the valuable species, mapping of the species habitat and detailed inventory of prioritized species are being practiced for NTFPs assessment in some countries. The methodologies are different for different species according to the types of NTFPs and their parts of use like rhizomes, leaves, barks, whole plants, fruits, flowers etc. These existing methods need to be reviewed and considered as a basis for developing better methodologies. In addition, relevant biodiversity attributes can be defined in cooperation with other organizations. Also rattan and bamboo information from NFI 1989–96 could be utilized in this context.

3 Development of Final Forest Inventory Design

3.1 Refinement of old maps and statistics

The free Global Canopy Height 2020 and Sentinel imageries provides a good opportunity to update previous forest cover maps. There are methodologies, which utilize segmentation techniques and attributes of old maps to generate spatially improved maps (Figure 13).

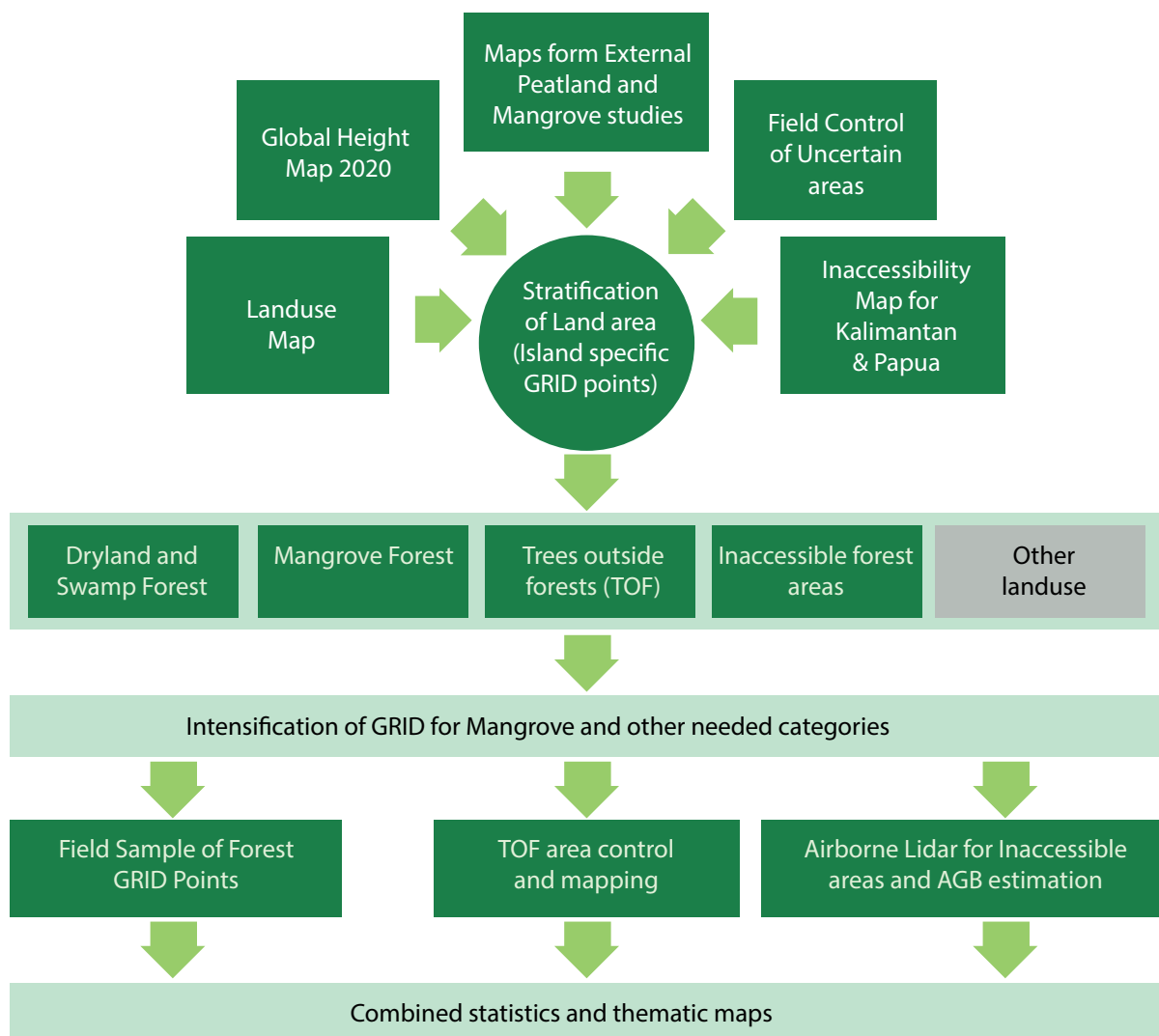


Figure 13. Flowchart of inventory process.

For peatland areas, there is already developed capacity to use tools (SEPAL and EX-ACT) that allow technically solid monitoring of peatlands. Also, the report of Krisnawati et al. (2015) represent the annual sum of all GHG emissions and removals presented as CO₂-eq. Dataset and methodologies developed in various previous projects can be utilised when final stratification is compiled.

The resources were limited when previous NFI results were calculated. The NFI 1989–96 database needs to be utilized fully for monitoring purposes. Also, FAO compatible results need to be derived from plot specific canopy cover observations.

3.2 Development of biometric models

New variables (e.g., biomass, carbon and growth) to be reported in the NFI may need specific localized biometric models so that sample-based information can be interpolated over all areas using remote sensing information. The NFI field sample plots are the primary source for biomass, carbon data measurements and modelling. The modelling of growth needs permanent sample plots information and requires long term commitment in data collection. Biodiversity and landscape parameters related methods and sampling schemes need to be developed in Indonesia. If feasible habitat mapping could be done using inventory data as well as supplementary secondary information. The mapping would be a valuable tool and it provides reference material for biodiversity, wildlife and plant surveys.

Following models need to be constructed:

- canopy cover models⁶,
- local calibration of biomass models,
- relation between biomass and carbon components,
- relation between biomass and ALS plots,
- relation between carbon and ALS plots,
- biodiversity index using plots data,
- plant diversity model based on plots and sample areas,
- plant diversity and habitats,
- remote sensing data and landscape livelihood.

3.3 Development of rapid change monitoring procedure

In addition to an NFI baseline survey, change monitoring is a key element of a National Forest Monitoring System (NFMS). It is necessary to improve observations of land and forest cover changes and to get a better understanding of mechanisms on linkages between deforestation and emissions. Changes in forest area, emissions and removals of greenhouse gasses over time can be assessed on lands drawn into the accounting framework using a combination of direct measurements, activity data, and models based on acceptable statistical principles, vegetation, land-use and land management surveys, forest inventories, remote sensing techniques, soil sampling and ecological surveys.

Spatial resolution has important cost and accuracy implications. In the Marrakesh Accords, the spatial resolution of forest is a minimum area of between 0.05 and 1.0 hectare, and revegetation is a minimum area of 0.05 hectare. If a small minimum resolvable area is used, the amount of task and cost of assessment can be high. With a coarse spatial resolution, the data demand can be modest but significant areas subject to an activity may be lost from the accounting system. How forest degradation and devegetation of other vegetation types are defined will determine the scale of effect that is captured within the accounting framework. What is considered forest degradation or devegetation of other vegetation types and the degree of change, loss or reduction that is required (e.g. 5% loss of biomass, 10% loss of carbon stocks, or 25% reduction in the supply of benefits) together with the spatial scale of estimation (0.05 or 1.0 hectare) will all determine how much affected land is drawn into the accounting system.

6 Canopy cover models are needed if ALS materials are available (see e.g. Korhonen *et al.*, 2015).

3.4 Practical tests and field surveys

A pilot inventory will be essential to test and refine the final system. The following goals can be set to this pilot:

- to compare nested circular plots and rectangular plots (partially on the top of old NFI PSPs), in order to obtain experience and time consumption data of outlined sampling designs,
- to practice the usage of new tools, such as Vertex 4, GPS field navigation and data capturing using tablets, Open Foris tools, and spherical densiometer (or equal device in canopy cover or closure assessments),
- to test applicability of very-high resolution images (e.g. NICFI Planet Labs images) for analyzing accessibility and land cover on the targeted sample sites,

3.5 Development of NFI Information System

There are needs for modern integrated database software for the forest inventory data input, processing and reporting. The software will be used in the central and regional offices, and also in the field. The calculation system should preferably also provide estimates for annual growth. Objectives of this task could be following:

1. To tailor or develop as software which can be used for field data result calculations. The proposed software should be running in a multi-user environment, e.g. in the cloud.
2. The system should allow to utilize sample plot data directly in remote sensing -based inventory data processing.

The data requirements and database structure need to be planned in these systems. The structure of the current tally sheets and the codes need to be surveyed. The FAO supported systems are one option in the final NFI implementation phase.

4 Acknowledgements

This proposal for NFI inventory design for Indonesia is the combined product of the efforts of a group of experts and institutions. Technical backstopping was provided by Adam Gerrand and Marco Piazza, and all possible help by Ibu Rubeta Andriani and Ibu Daya Sebayang (all FAO). Gael Sola (FAO) has only committed valuable assistance for this planning process. The report compilers would like to extend their gratitude to all contributors to the development of this NFI design plan and in particular the below mentioned.

Great thanks to all MoEF staff who guided the NFI planning process, participated in the consultation meetings and in the process of collecting and compiling information, particularly Dr. Belinda Arunarwati Margono, Ibu Nurhayati, and Mr. Judin Purwanto. Also special thanks to the FAO National Experts Dr. Tatang Tiryana, Dr. Solichin Manuri, Mr. Abdul Mukti, Mr. Edwin Nugroho for their important roles in providing knowledge, data and background information about Indonesia forests, spatial data, human resources, and costs.

Finally, a big word of thanks to FAO Country Representative Mr. Rajendra Aryal who played a crucial role in enabling administrative processes to go smoothly to the goal and the FAO-Indonesia office staff for their valuable support in organizing consultancy missions and logistics.

5 References

- Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Bafort, W. & Heinzen, D.F. (1994). Satellite Inventory of Minnesota Forest Resources. *Photogrammetric Engineering and Remote Sensing*, 60: 287-298.
- Brown, I.F., Martinelli, L.A., Thomas, W.W., Moreira, M.Z., Ferreira, C.A.C., & Victoria, R.A. (1995). Uncertainty in the Biomass of Amazonian Forests - an Example from Rondonia, Brazil. *Forest Ecology and Management* 75: 175–189.
- Clark, D.B. & Clark, D.A. (2000). Landscape-scale variation in forest structure and biomass in a tropical rainforest. *Forest Ecology and Management* 137:185-198.
- Cochran, W. (1977). *Sampling techniques*, 3rd ed. John Wiley & Sons, New York. 428 p.
- Dubayah, R., Armston, J., Healey, S.P., Bruening, J. M., Patterson, P.L., Kellner, J. R., Duncanson, L., Saarela, S., Ståhl, G., Yang, Z., Tang, H., Blair, J. B., Fatoyinbo, L., Goetz, S., Hancock, S., Hansen, M., Hofton, M., Hurtt, G. and Luthcke, S. 2022. GEDI launches a new era of biomass inference from space. *Environ. Res. Lett.* 17 095001. DOI 10.1088/1748-9326/ac8694. (https://www.researchgate.net/publication/362464680_GEDI_launches_a_new_era_of_biomass_inference_from_space)
- FAO. (2007). Brief on National Forest Inventory NFI Indonesia. MAR-SFM Working Paper 18. <https://www.fao.org/3/ap186e/ap186e.pdf>
- FAO (2016). Land Cover Classification System: Classification concepts. <https://www.fao.org/3/i5232e/i5232e.pdf>
- Fehrmann, L., Kleinn, C., Magdon, P. & Pérez-Cruzado, C. (2017). Definition of a minimum standard for forest management inventories on KPH level. FORCLIME Technical Cooperation between MoEF and GIZ. [https://www.forclime.org/documents/Books/Minimum%20standard_final%20%20finish%20Ready%20print%20\(Ingggris\)%20Lowrest.pdf](https://www.forclime.org/documents/Books/Minimum%20standard_final%20%20finish%20Ready%20print%20(Ingggris)%20Lowrest.pdf)
- Finley, A.O., Banerjee, S., MacFarl, D.W. 2011. A Hierarchical Model for Quantifying Forest Variables Over Large Heterogeneous Landscapes. With Uncertain Forest Areas. *Journal of the American Statistical Association* , March 2011, Vol. 106, No. 493. (March 2011), pp. 31-48
- GOFC-GOLD (2016). Sourcebook. A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation. <http://www.gofcgold.wur.nl/redd/>
- Hou, Z., Xu, Q., Hartikainen, S., Anttila, P., Packalen, P., Maltamo, M. & Tokola, T. (2015). Impact of Plot Size and Spatial Pattern of Forest Attributes on Sampling Efficacy. *Forest Science* 61(5): 847-860(14).
- IPCC. (2003). Good Practice Guidance for Land Use, Land-Use Change and Forestry. https://www.ipcc-nggip.iges.or.jp/public/gpoglulucf/gpoglulucf_files/GPG_LULUCF_FULL.pdf
- Kershaw Jr., J.A., Ducey, M.J., Beers, T.W. & Husch, B. (2016). *Forest Mensuration*, 5th ed. John Wiley & Sons Ltd., UK.
- Kleinn, C. (1996). Ein Vergleich der Effizienz von verschiedenen Clusterformen in forstlichen Grossrauminventuren. *Forstw Cbl.* 115: 378–390.
- Köhl, M. (1990). Double sampling in the Swiss National Forest Inventory. Results of a Pilot Study. Proceedings from SNS/IUFRO workshop in Umeå 26-28 Feb. 1990. Remote Sensing Laboratory, Swedish University of Agricultural Sciences, Report 4: 76-83.
- Köhl, M. & Scott, C. T. (1992). Survey Planning for National Forest Inventories. Proceedings of the Conference: Integrating Forest Information over Space and Time. IUFRO Conference, January 13-17 1992, Canberra, Australia.
- Korhonen, L., Ali-Sisto, D. & Tokola, T. (2015). Tropical forest canopy cover estimation using satellite imagery and airborne lidar reference data. *Silva Fennica* vol. 49 no. 5 article id 1405. <https://doi.org/10.14214/sf.1405>

- Korhonen, L. Packalen, P. & Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote sensing of environment*, 2017, 195: 259-274.
- Krisnawati, H., Imanuddin R., Adinugroho, W. C. Hutabarat, S. (2015). National Inventory of Greenhouse Gas Emissions and Removals on Indonesia's Forests and Peatlands. H. Krisnawati. Bogor, Indonesia, INCAS MoEF Research, Development and Innovation Agency of the Ministry of Environment and Forestry.
- Lang, N., Jetz, W., Schindler, K. & Wegner, J. D. (2022). A high-resolution canopy height model of the Earth. arXiv preprint arXiv:2204.08322. (Data available for download: https://share.phys.ethz.ch/~pf/nlangdata/ETH_GlobalCanopyHeight_10m_2020_version1/)
- Magnussen, S., Russo, G. 2012. Uncertainty in photo-interpreted forest inventory variables and effects on estimates of error in Canada's National Forest Inventory. *The Forestry Chronicle*. 1 January 2012. <https://doi.org/10.5558/tfc2012-080>
- Matern, B. 1960. Spatial variation. *Medd. Statens Skogsf. Inst.* 49(5). Also appeared as number 36 of Lecture Notes in Statistics. Springer-Verlag, New York, 1986.
- Matern, B. (1981). Funderingar om provträd i riksskogstaxeringen. *Inst. f. Skogstaxering, SLU, Umeå. NUTAX-Rapport Nr 8.*
- MoEF. (2015). National Forest Reference Emission Level for Deforestation and Forest Degradation: In the Context of Decision 1/CP.16 para 70 UNFCCC (Encourages developing country parties to contribute to mitigation actions in the forest sector). Ministry of Environment and Forestry (MoEF), Republic of Indonesia.
- Nanjaya, E.K., Rusolono, T. & Tiryan, T. (2020). Penentuan Ukuran Plot Contoh Optimal Untuk Pendugaan Luas Bidang Dasar dan Biomassa Tegakan. (Determination of Optimal Sample Plot Size for Estimating Stand Basal Area and Biomass). *Penelitian Hutan Dan Konservasi Alam*, 17: 65-77.
- Pelz, D.R. & Cunia, T. (ed.) (1985). National Forest Inventories in Europe, Proceedings, *Mitteilungen der Abteilung Forstliche Biometrie der Universität Freiburg*, Nr. I.
- Ranneby, B. (1981a). Provytestorlekens betydelse vid skogsinventeringar. *Inst. f. Skogstaxering, SLU, Umeå. NUTAX Rapport Nr 4.*
- Ranneby, B. (1981b). Den topografiska variationen inom olika områden. En redovisning av skattade korrelationsfunktioner. *Inst. f. Skogstaxering, SLU, Umeå. NUTAX Rapport Nr 5.*
- Schreuder, H. T., Gregoire, T. G. & Wood, G. B. (1993). *Sampling Methods for Multiresource Forest Inventory*. John Wiley & Sons, Apr 16, Nature. 464 p.
- Slik, J., G. Paoli, K. McGuire, I. Amaral, J. Barroso, M. Bastian, L. Blanc, F. Bongers, P. Boundja, C. Clark, M. Collins, G. Dauby, Y. Ding, J.-L. Doucet, E. Eler, L. Ferreira, O. Forshed, G. Fredriksson, J.-F. Gillet, D. Harris, M. Leal, Y. Laumonier, Y. Malhi, A. Mansor, E. Martin, K. Miyamoto, A. Araujo-Murakami, H. Nagamasu, R. Nilus, E. Nurtjahya, Á. Oliveira, O. Onrizal, A. Parada-Gutierrez, A. Permana, L. Poorter, J. Poulsen, H. Ramirez-Angulo, J. Reitsma, F. Rovero, A. Rozak, D. Sheil, J. Silva-Espejo, M. Silveira, W. Spironelo, H. ter Steege, T. Stevart, G.E. Navarro-Aguilar, T. Sunderland, E. Suzuki, J. Tang, I. Theilade, G. van der Heijden, J. van Valkenburg, T. Van Do, E. Vilanova, V. Vos, S. Wich, H. Wöll, T. Yoneda, R. Zang, M.-G. Zhang, N. Zweifel. (2013). Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics. *Glob. Ecol. Biogeogr.*, 22: 1261–1271.
- Sutter, H. (1990). Field Instructions for the National Forest Inventory Indonesia. FAO. UTF/INS/066/INS NFI Field document No. 6. 41 p.
- Tiryan, T. (2022). Needs Assessment for designing the NFI system of Indonesia. Project paper, FAO Indonesia.
- Tokola, T. (2015). Remote Sensing Concepts and their Applicability in REDD+ Monitoring. *Current Forestry Reports*, 1 (4), 252–260.
- Tokola, T., Pitkänen, J., Partinen, S. & Muinonen, E. (1996). Point Accuracy of a Non-Parametric Method in Estimation of Forest Characteristics with Different Satellite Materials. *International Journal of Remote Sensing* 17(12): 2333–2351.
- Varjo, J. (1997). Change Detection and Controlling Forest Information Using Multi-temporal Landsat TM Imagery. *Acta Forestalia Fennica* 258. 64 p.
- Zhang, S., Vega C., Deleuze C., Durrieu S., Barbillon P., Bouriaud O. Renaud J-P. 2022. Modelling forest volume with small area estimation of forest inventory using GEDI footprints as auxiliary information. *International Journal of Applied Earth Observation and Geoinformation*, Volume 114, November 2022, 103072. <https://doi.org/10.1016/j.jag.2022.103072>

Appendix

Appendix 1. Review of alternative data and methods

1.1 Sampling strategies

When a large-area forest data collection is planned, we typically complete the following steps:

1. Study of target population: incl. Landscape patterns and Forest structure.
2. Comparison of remote sensing materials for stratification.
3. Planning of suitable fieldwork design for different variables and forest types.
4. Compare alternatives for updating the procedure.

One of the best examples of the forest data collection planning process is the Swedish NFI. As a basis for the statistical design of the survey, a geostatistical analysis has been used to determine the variation within areas, the importance of the size of the sample plot, the time required and the economic practicability of the available resources. The analysis has resulted in a division of the country into 5 regions, the designing of survey tracts, a weighting between permanent and temporary survey tracts and a standard size of sample plot (Matern, 1960; Matern, 1981; Ranney, 1981a; Ranney, 1981b). In addition, variograms have been used to describe variations in land use, forest volume and topography (Matern, 1960; Ranney, 1981b), and these spatial functions have been used to define an effective layout for the survey tracts. Variograms are describing how neighboring areas are correlated and how variation of forest characteristics are changing according to distance from specific points.

In statistical sampling, the aim of stratification is to divide the population into smaller subpopulations so that different sampling intensity or even completely design can be used in the different strata. Stratification criteria can be based on importance of the strata, variability inside the strata or assessment costs. To avoid too small samples in an NFI, the number of strata cannot be very high.

One logical stratification criterion in Indonesia is the archipelago, its islands. These are stable areas in terms of their geographic position (which is a weakness with administrative borders that tend to change over the time). However, there is still some geographical variation within some of these islands, and some areas are more difficult to access than others: for example, mountainous areas (i.e. 1000 m asl.) can be very difficult to access. In addition, there are some security issues in Aceh, Central Sulawesi and Papua and these may hinder accessing some forest areas. Therefore, stratification based on geographic features (islands, mangrove forests, high mountain areas and restricted areas) can be considered.

Another widely used stratification criterion is the measuring cost of the plots (or plot clusters). Plots or clusters with high measuring costs should have lower sampling intensity than the low-cost plots or clusters. The cost relates to distance from settlement and roads and accessibility. Therefore, it is a very useful stratification criteria when assessment of changes is of interest: it is realistic to assume that most changes take place in areas near to settlements and roads (where the assessment cost would be low and sampling intensity high). However, stratification by measuring costs is sensitive to quality of the cost assessment: maps need to be up-to-date to correctly show locations of all roads and settlements.

Forest inventory systems based on remote sensing and their implementations can be categorized in the following way:

1. Remote-sensing-oriented systems; the main attention is given to the identification of different forest classes, including species classes and volume categories, meant for thematic map production purposes. Ground truthing is performed from subjectively chosen training areas with some field observations.
2. Field-data-oriented systems; the calculation of inventory results is based on field sample plots and the inventory system is established according to a pre-designed sampling frame. Remote sensing data is used as auxiliary information mainly in two-phase sampling schema (e.g. Köhl, 1992) or in calibration approaches (e.g. Bauer *et al.*, 1997).
3. Updating-oriented-systems; the main interest is to fulfill the information needs of the database and control the quality of data. The base information is mostly collected with more accurate inventory systems, and remote sensing data are used for monitoring purposes and to allocate field checking (Varjo, 1997). The existing data can be utilized as auxiliary data in the planning of a data collection procedure.

The main methods for the estimation of forest characteristics are stratification (or classification) and regression analysis. All estimates derived using these methods are eventually based on field data and remote sensing data is generally used to expand the data by interpolation over non-sampled areas. In the case of the regression approach, the variables of interest are modeled separately. In the stratification and non-parametric estimation (e.g., k-nearest neighbor estimation method, kNN) approaches, however, several variables can be estimated simultaneously.

In forest inventory problems, the size of the plots to be measured needs also to be decided in addition to the selection of the number of plots. The larger the plots are, the more time-consuming and expensive it is to measure them, and obviously the number of sample plots that can be measured with a given budget is larger when the plots are small. On the other hand, the variation among plots in the population, variation, diminishes as the plot size increases. Clusters of smaller subplots (or combined plots) have been used on many occasions instead of single plots, typically in large-area surveys such as national inventories. A cluster plot typically consists of small circular or rectangular plots that form a geometrical figure such as a triangle or rectangle. There are two benefits entailed in the use of clusters. First, the location and layout of a cluster of several small plots are faster and more accurate than the measuring of large single plots. In addition, the coefficient of variation is smaller than for single plots of the same total area.

The optimal plot size thus depends on both the measurement costs and the observed variation. It can be shown that the spatial pattern of forests affects the optimal plot size. The size of the plot also determines the inclusion area for each tree. When circular plots of radius " r " are used, the inclusion area is a circle also of radius " r " centered on the tree. In other words, a certain tree will be measured if the sample plot is located in its inclusion area.

Increment estimation is normally based on increment borings and height increment measurements performed on sample trees (height increment models in the case of broad-leaved trees). In principle, this method corresponds to the use of permanent plots for increment estimation but produces estimates with a lower standard error than the method which uses volume differences on permanent plots, due to the fact that errors in diameter measurements are usually greater than errors in the measurement of diameter increment cores. Another possibility is that a certain proportion of the plots in the entire country is being measured each year. Thus country-level estimates can be updated annually and regional-level estimates within 2–3 years of the start of the survey.

Structure of forest areas and forest types

The relative spatial distribution of forests and trees varies, because of changing land-use practices, differences in the fertility of the soil, and the hydrology, competition, and size distribution of trees (e.g. Hou *et al.*, 2015). On a regional level, both correlation and autocorrelation functions have been employed to study large structures of forest area, and volume of forests. For instance in Scandinavian forests, there is a slightly increasing autocorrelation almost until a 200 m distance when larger distances are applied. This information is used to estimate the optimal distance between inventory tracts, the overall shape of the tract, and the distance between sample plots within tracts. In order to obtain a sufficient sampling set-up for the entire nation of Sweden, the country is divided into five regions with marginally diverse correlation functions for different variables. The final sampling design is based on their spatial characteristics and other practical considerations.

Many forestry variables are spatially sparse and scattered. This is often the case when one is assessing coarse woody debris in managed forests, or surveying threatened species. The spatial description of sparse populations can also be problematic. On the landscape level, information about the spatial distribution of different key habitats and areas with a high ecological value has also been used to assess the probability of existing rare species. Field data about indicator species and remote sensing data about landscape features are valuable *a priori* information for estimating the presence/absence probability and for stratifying areas of interest.

Canopy cover

In its forest resource assessments, FAO uses forest definition of a minimum cover of 10%, height of vegetation 5 m, and area of 0.5 ha. However, the FAO approach of a single worldwide value excludes variability in ecological conditions and differing perceptions of forests. For the purpose of the Kyoto Protocol, it was determined through the Marrakesh Accords that Parties should select a single value of the crown area, tree height, and area to define forests within their national boundaries. The selection must be from within the following ranges, with the understanding that young stands that have not yet reached the necessary cover or height are included as forest:

- Minimum forest area: 0.05 to 1 ha,
- Potential to reach a minimum height at maturity *in situ* of 2–5 m,
- Minimum tree crown cover (or equivalent stocking level): 10% to 30%.

Under this definition a forest can contain anything from 10% up to 100% tree cover; it is only when cover falls below the minimum crown cover as designated by a given country that land is classified as non-forest. However, if this is only a temporary change, such as for timber harvest with regeneration expected, the land remains in the forest classification. The specific definition chosen will have implications on where the boundaries between deforestation and degradation occur. The definition of forests offers some flexibility for countries when designing a monitoring plan because analysis of remote sensing data can adapt to different minimum tree crown cover and minimum forest area thresholds. However, consistency in forest classifications for all REDD activities is critical for integrating different types of information including remote sensing analysis.

Typically, the FAO/LCCS manual recommends to apply the following categories for land cover mapping (FAO 2000):

Very Dense Forest	All lands with tree cover of canopy density of 70% and above
Moderately Dense Forest	All lands with tree cover of canopy density between 40 % and 70 % above
Open Forest	All lands with tree cover of canopy density between 10 – 40 %.
Scrub	All forest lands with poor tree growth mainly of small or stunted trees having canopy density less than 10 percent.
High grass	Grassland with grasses above 1 m.
Non-forest	Any area not included in the above classes.

However, categories can be adjusted according to the typical structure of vegetation in the region. In Indonesia, land use maps utilize 30 % crown cover criteria and 6.25 ha minimum area for forest definition. This definition is largely affected by properties of previous Landsat generation of images. In this material it was very difficult to discriminate between forest and scrubland in tropical areas and the pixel size was rather large at that time.

1.2 Review of remote sensing and vegetation mapping alternatives

There are many alternative Remote Sensing materials for regional biomass and carbon inventory. The selection of satellite imagery data is not a very easy task. Traditionally 20–30 m ground resolution images are used in national level inventories. However, it has been found in many studies that separation between scrub and woody vegetation is difficult from such material. There is essentially a need for 2.5–10 m multispectral resolution because the texture of the vegetation is needed for separation of trees and other green vegetation. The availability of data brings additional difficulties and most probably the final set of remote sensing data is a combination of different materials (e.g. Tokola *et al.* 1996, Tokola 2015, Korhonen *et al.* 2017). There are many data sharing concepts available (e.g. NICFI Planet Labs images). Through the Norway's International Climate & Forests Initiative (NICFI), users can now access Planet Lab's high-resolution, analysis-ready mosaics of the world's tropics in order to help reduce and reverse the loss of tropical forests, combat climate change, conserve biodiversity, and facilitate sustainable development for non-commercial uses

In the traditional national inventory, biomass estimates are based on field measurements of the number of trees per ha combined with the diameter and height of each tree. To update such measurements is both time-consuming and expensive. Costs may be reduced by updating only a fraction of the total area each year, but this will obviously lead to global estimates always being somewhat outdated. New technology is being introduced for this purpose. By measuring the time of laser light emitted from equipment (LiDAR, ALS) mounted on aircraft to be reflected from the ground and the canopy of vegetation, one may estimate biomass density.

The pilot tests are normally implemented in those conditions where the estimation performance of specific remote sensing material (R^2) is not very well known. The coefficient of determination (R^2) is used in the context of statistical models whose main purpose is the prediction of future outcomes based on other related information, and it is the proportion of variability in a data set that is accounted for by the statistical model. It provides a measure of how well future outcomes are likely to be predicted by the model. In this report, analysis is made for different R^2 levels and remote sensing material-specific coefficients have been estimated using literature and the best knowledge gained in other countries. The capacity of very high-resolution images has been assumed to be close to aerial photographs. Airborne lidar (ALS) provides the same biomass estimation capacity than field surveys, but species-specific information and other ground characteristics are missing.

1.3 Potential materials for stratification of vegetation in NFI

When R^2 value between field data and remote sensing data is increasing the need for fieldwork is decreasing heavily. The size of the field sample plot is also important. If we have too small field sample plots, the variation between plots is very high. The inclusion probability for small areas for large trees is so low that relatively large plots are required. However, it is recommended that the whole plot should fall in the same forest type, so that remote sensing procedures can utilize plot information directly as a training area.

The cost and quality of remote sensing are normally connected to each other. The medium resolution satellite data (e.g. Sentinel and Landsat) are quite cheap or free for area units and we can easily cover the entire country at a low cost. VHR satellite data and ALS data are normally utilized on a sample basis. Depending on the autocorrelation of forest population, a certain type of information can be collected from a single sample unit.

When the cost of field measurements is considered, the difference between small and large plots is quite small within different remote sensing materials. The field data cost can be added to image data costs and the total cost of the inventory system can be estimated. Normally, we need a whole country covering the dataset and VHR/ALS dataset with proper field sampling. Thus, the final system is a combination of full coverage, remote sensing sample and field plots. So, according to the current price assumptions, the most cost-efficient way is to use Sentinel data, optional laser scanning and field data. In this case, field plots are planned for forest areas only.

Table 9 and Table 10 demonstrate characteristics of potential sensors for acquiring suitable data for detailed stratification of forest areas. MODIS data has been used in large area inventories, but its spatial resolution is poor for delineation of forest areas. Landsat and Sentinel data are principally potential alternatives. Detailed LiDAR data is a rather expensive data source, but spaceborne Global Ecosystem Dynamics Investigation (GEDI) data can provide direct measurements from its sample line assessments.

There are four ready compiled Global product which could be used in biomass stratification:

1. Global 30m Landsat Tree Canopy 2015.
2. Global Forest Canopy Height, 2019, University of Maryland (UMD GLAD).
3. Global Canopy Height 2020, EcoVision Lab, Photogrammetry and Remote Sensing, ETH Zürich.
4. GEDI L4A Footprint Level Aboveground Biomass Density, Version 2.1, NASA.

Table 9. Review of MODIS, Landsat and Sentinel Data.

Description	Modis			Landsat 8			Sentind-2		
	Band	Wavelength	Resolution	Band	Wavelength	Resolution	Band	Wavelength	Resolution
Blue	Band 3	459-479	500	Band 2	452-512	30	Band 2	458-523	10
Green	Band 4	545-565	500	Band 3	533-590	30	Band 3	543-578	10
Red	Band 1	620-670	500	Band 4	636-673	30	Band 4	650-680	10
NIR	Band 2	841-876	500	Band 5	851-879	30	Band 8	785-899	10
SWIR-1	Band 6	1628-1652	500	Band 6	1566-1651	30	Band 11	1565-1655	20
SWIR-2	Band 7	2105-2155	500	Band 7	2107-2294	30	Band 12	2100-2280	20

Table 10. Selection of alternative LiDAR datasets.

LiDAR Platform	Sensor	Recording mode	Wavelength	Footprint size	Spatial extent
Spaceborne	GLAS (ICESat satellite)	Full waveform Sampling (1 footprint every 170 m along laser ground track)	532nm/ 1064 nm		Global ($\pm 86^\circ$ N/S latitude)
ATLAS (ICESat-2 satellite)	Photon-sensitive discrete returns	Profiling (1 footprint every 70 cm along laser ground track)	532 nm	~14 m	~70 m
GEDI (ISS)	Full waveform	Sampling (1 footprint every 60 m along ground track)	1064 nm	~25 m	Near-Global ($\pm 51.6^\circ$ N/S latitude), Global ($\pm 88^\circ$ N/S latitude)
Airborne	Overview of multiple conventional linear mode systems	Discrete returns Scanning or profiling	800 nm - 2500 nm (Near-Infrared range)	10-30 cm at 500-3000 m range	Regional to country scale
Land, Vegetation and Ice Sensor (LVIS)	Full waveform	Scanning	1064 nm	20-30 m	Regional to country scale Limited to 1 instrument operated by NASA
SPL100 (Leica/Hexagon)	Photonsensitive discrete returns	Scanning	532 nm	20-30 cm at 3000-4000 m range	Regional to country scale Potential canopy penetration and ground retrieval limitations under dense vegetation (Brown <i>et al.</i> , 2020)
UAS	Overview of multiple conventional systems	Discrete returns Scanning	700 nm	2.5-30 cm at 50-100 m range	Local

1.3.1 Global 30m Landsat Tree Canopy 2019

The Landsat Tree Canopy Cover⁷ layers contain estimates (as the percentage of horizontal ground) in each 30-m pixel covered by woody vegetation (greater than 5 meters in height). The product is derived from all seven bands of Landsat-5 TM and/or Landsat-7 ETM+.

Tree cover, the proportional, vertically projected area of vegetation (including leaves, stems, branches, etc.) of woody plants above a given height, affects terrestrial energy and water exchanges, photosynthesis and transpiration, net primary production, and carbon and nutrient fluxes. Tree cover also affects habitat quality and movements of wildlife, residential property value for humans, and other ecosystem services. The continuous classification scheme of the VCF product enables better depiction of land cover gradients than traditional discrete classification schemes. Importantly for detection and monitoring of forest changes (e.g., deforestation and degradation), tree cover provides a measurable attribute upon which to define forest cover and its changes.

7 <https://landsat.gsfc.nasa.gov/article/global-30m-landsat-tree-canopy-version-4-released/>

The processing methodology with this dataset is based on calculation of tree cover. One method is to use 10-m resolution, and then aggregate rasters up to 250-m resolution by averaging the values within the extent of each 250-m pixel. In pixels with steep underlying terrain, the varying ground elevation in large pixels can cause spurious detection of tree cover as LiDAR returns above 5 m height, so computing first cover in small, 10-m size pixels and then aggregating these to 250-m pixels can be used to avoid this effect.

Dataset Reference: Sexton, J. O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R.G. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS Vegetation Continuous Fields with LiDAR-based estimates of error. *International Journal of Digital Earth*, 130321031236007. doi:10.1080/17538947.2013.786146.

1.3.2 Global Forest Canopy Height 2019

A 30-m spatial resolution global forest canopy height map was developed through the integration of the GEDI LiDAR-based forest structure measurements and Landsat analysis-ready data time-series (Figure 14). NASA's GEDI is a spaceborne LiDAR instrument operating onboard the International Space Station since April 2019.

It provides footprint-based measurements of vegetation structure, including forest canopy height between 52°N and 52°S globally. The Global Land Analysis and Discover team at the University of Maryland (UMD GLAD) integrated the GEDI data available to date (April-October 2019) with the year 2019 Landsat analysis-ready time-series data (Landsat ARD). The GEDI RH95 (relative height at 95%) metric was used to calibrate the model.

Dataset Reference : P. Potapov, X. Li, A. Hernandez-Serna, A. Tyukavina, M.C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C.E. Silva, J. Armston, R. Dubayah, J. B. Blair, M. Hofton (2020) Mapping and monitoring global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, 112165. <https://doi.org/10.1016/j.rse.2020.112165>

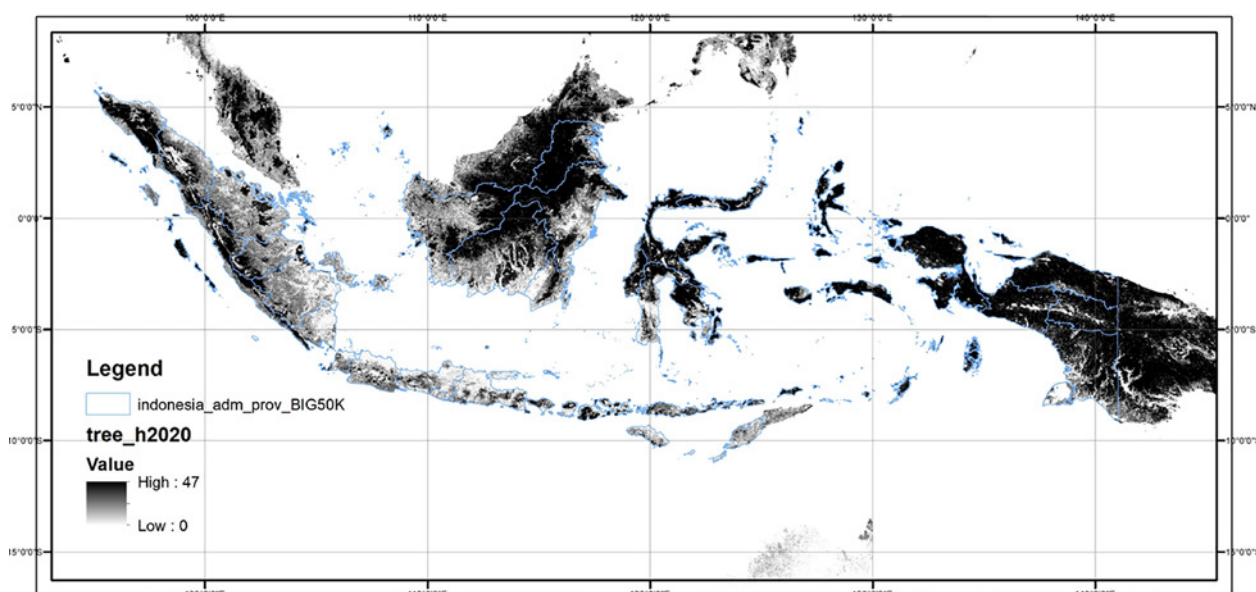


Figure 14. Example of canopy height model covering entire Indonesia, 2019. Ground resolution is 30 m.

1.3.3 Global Canopy Height 2020

This Global canopy top height dataset is for the year 2020 at a 10 m ground sampling distance (Figure 15). A probabilistic deep learning model has been developed to retrieve canopy top height from Sentinel-2 images anywhere on Earth. This model, an ensemble of convolutional neural networks (CNN) is trained with sparse supervision from GEDI derived canopy top height data (i.e. estimated RH98 from Lang *et al.*, 2022). Furthermore, the predictive uncertainty of these dense estimates is quantified. That approach reduces the saturation effect commonly encountered when estimating canopy height from optical satellite images allowing to resolve tall canopies likely containing high carbon stocks. The global wall-to-wall map is based on Sentinel-2 images taken between May and September 2020.

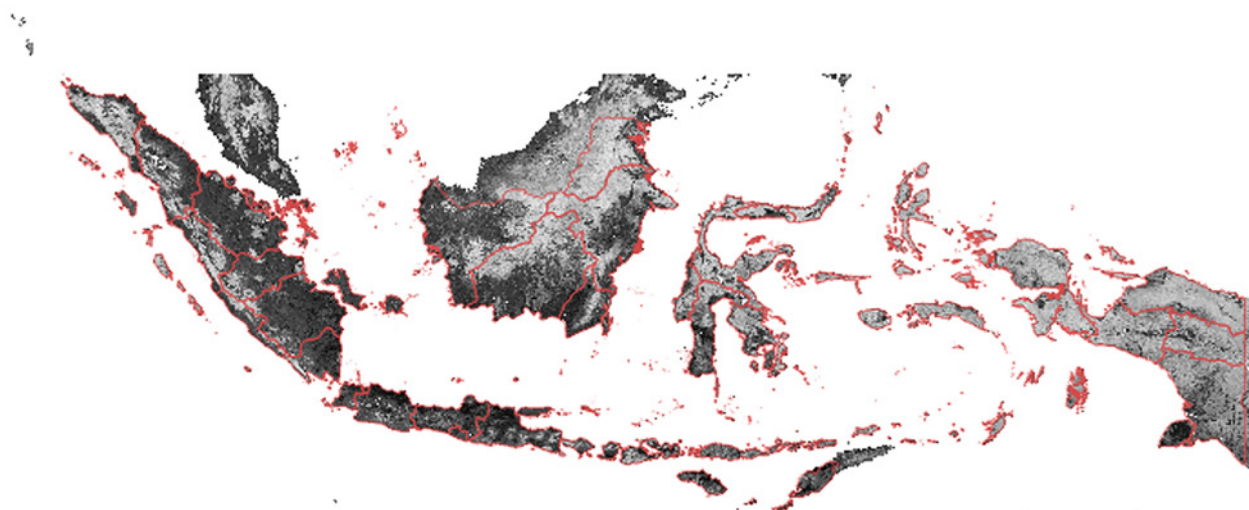


Figure 15. Example of canopy height model covering entire Indonesia, 2020.
Ground resolution is 10 m.

1.3.4 GEDI L4A Footprint Level Aboveground Biomass Density

This dataset contains Global Ecosystem Dynamics Investigation (GEDI) Level 4A (L4A). Version 2 predictions of the aboveground biomass density (AGBD; in Mg/ha, equal to tons/ha) and estimates of the prediction standard error within each sampled geolocated laser footprint⁸ (Figure 16). The footprints are located within the global latitude band observed by the International Space Station (ISS), nominally 51.6 degrees N and S and reported for the period 2019-04-18 to 2022-06-09. The GEDI instrument consists of three lasers producing a total of eight beam ground transects, which instantaneously sample eight ~25 m footprints spaced approximately every 60 m along-track. The GEDI beam transects are spaced approximately 600 m apart on the Earth's surface in the cross-track direction, for an across-track width of approx. 4.2 km.

1.3.5 Stratification and criteria

Following criteria can be used in selection materials for stratification:

1. Datum, up-to-date situation,
2. Availability and costs,
3. Spatial resolution and completeness,
4. Reliability of datasets for desired parameters.

⁸ If the GEDI data is used, it is important first to filter high quality observations from raw data (using e.g. these parameters: L2_Quality_Flag = 1, Degrade_Flag = 1 and Sensitivity > 95).

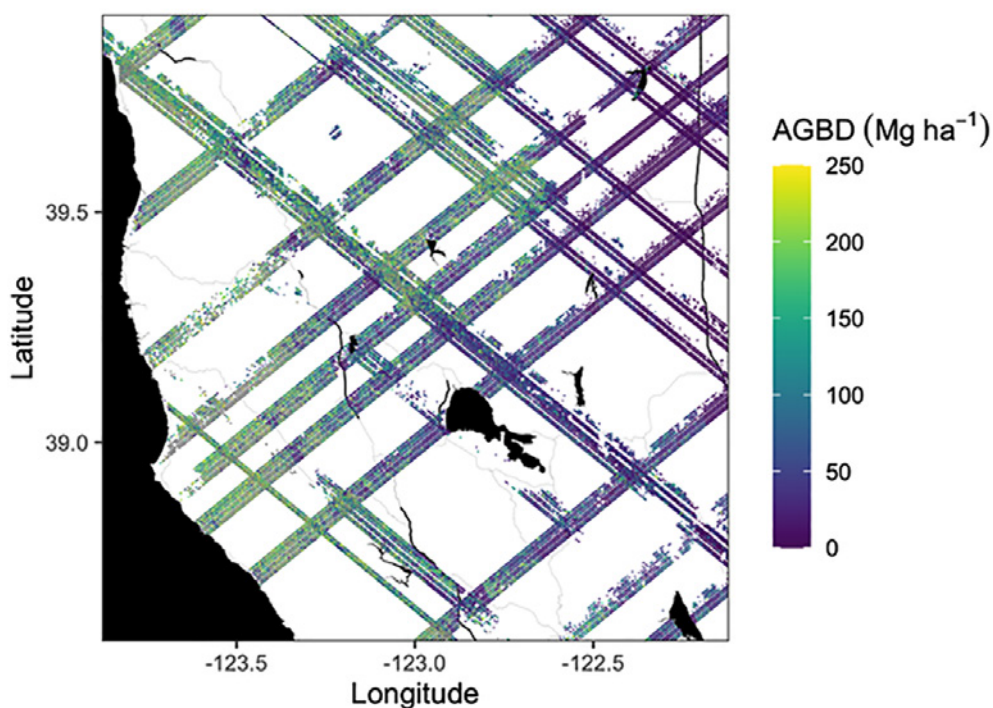


Figure 16. Example of GEDI above-ground biomass data.

The crown cover percentages can be defined for VHR based tree height data. Additionally, the crown cover percentage can be estimated for every class after getting proper field data. The resolution of the VHR data is closer to ground reference data. Tree cover can also be calculated from canopy height values within a 30-m neighborhood. In accordance with the International GeosphereBiosphere definition of forests, in the field, we specify the height criterion (of forest) to be 5 meters.

However, the ETH satellite product typically overestimates canopy heights and the criteria was set a threshold between 10–25 m (Figure 17). The following calculation chain was applied to detect tree coverage at 10 m resolution: 3x3 pixels window was used to calculate sum of sufficient amount of high-enough tree pixels for forest cover (> 10–20%). Frequency of forest pixels in 3x3 and 5x5 pixels window were applied in the following way:

- 0 – 1 pixel: 0 %
- 2 – 3 pixels: 20–40 %
- 4 – 6 pixels: 40–70%
- 7 – 9 pixels: > 70%

This procedure needs to be further tested in different Islands. Small image windows do not have a large effect on the interpretation results; these affect only clearly in borders of forest areas. In Indonesian data, canopy cover was rather dense e.g. in Kalimantan and these dense vegetated and forested blocks could be found applying canopy height thresholds. For further applications while planning to utilize radar data, there is still a need to compile prediction models for canopy coverage (e.g. Korhonen *et al.*, 2015)

According to VHR data and visual inspection, the initial stratification can be done using the following classes:

1. Islands (Sumatra, Jawa, Bali Nusa, Kalimantan, Sulawesi, Maluku, Papua),
2. Land cover map (baseline land cover),
3. ETH forest area, forest cover and canopy height maps (biomass variation), and FAO canopy cover,
4. GEDI sample lines (to test spatial correlation within forest).

The stratification map could contain four classes (i.e., dense forest, sparse forest, non-forest, and uncertain forest). An above-ground biomass map can also be compiled for testing alternative field sampling designs. On the top of this map, various cluster designs and densities can be generated and examined. In addition, overlapping of old NFI sampling sites can be still considered.

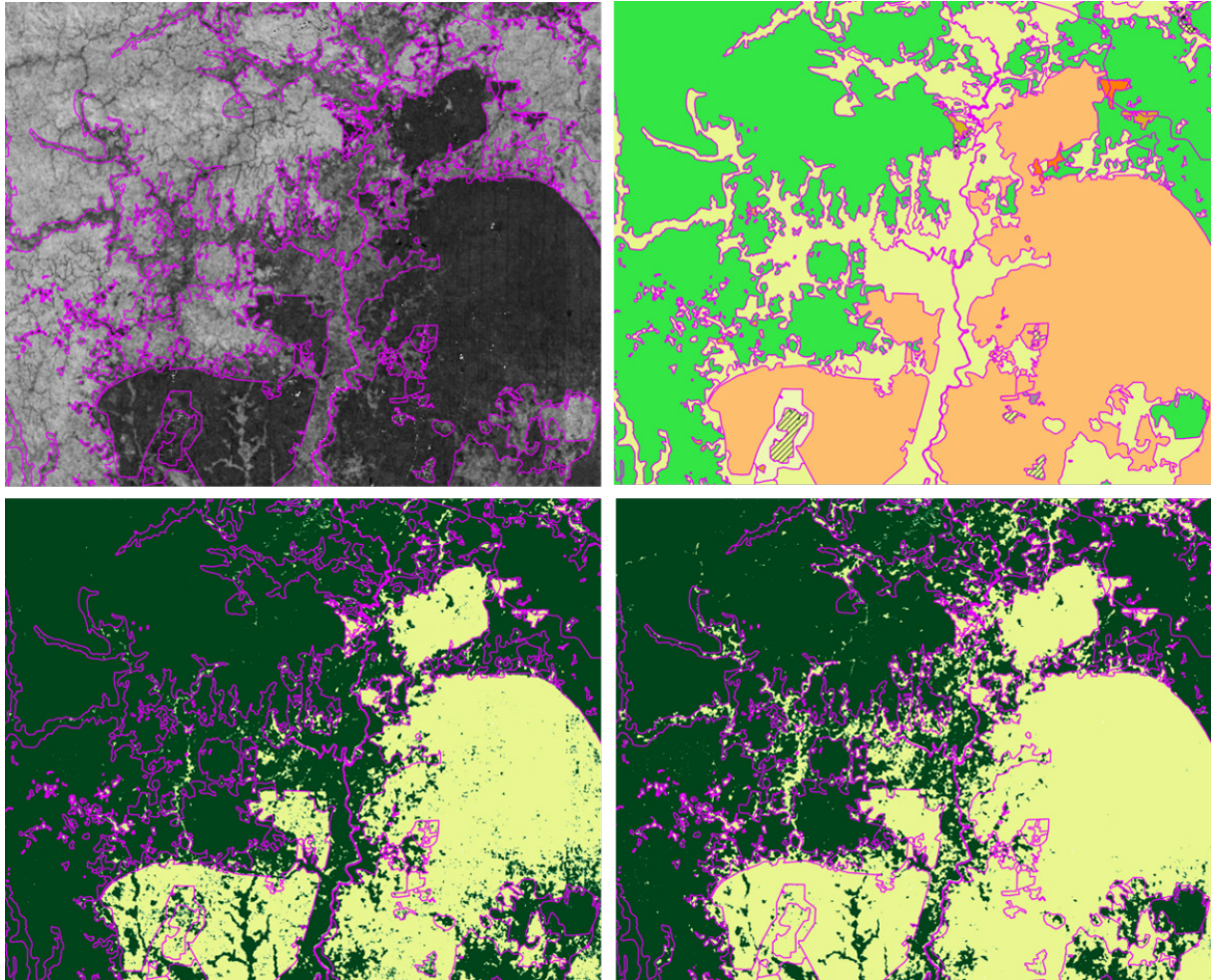


Figure 17. ETH Canopy Height image (top left), Indonesian Land cover map (top right), 15 m height threshold forest map (low left) and 20 m height threshold forest map (low right) describe status of land cover.

The final stratification scheme needs to be tested and developed for different Islands, when interpretation of recent satellite image material is done and some control of “uncertain” category is completed. It is important that ecological characteristics of different vegetation areas are controlled, when final forest-non-forest stratification is carried out.

Appendix 2. Forest Structure and NFI Design Analysis

2.1 Analysis of Field Plot Data

For a realistic forest inventory design we need to get information about forest population and its variation (regarding the variables of interest). The previous National Forest Inventory provides large countrywide data about forests and it is an excellent source for this purpose. The field data for the NFI system was obtained from measurements of cluster-plots, consisting of Temporary Sample Plots (TSPs) and Permanent Sample Plots (PSPs), which were established systematically in the grid of 20 km x 20 km to cover forested lands in all provinces (excluding Java Island) below 1000 m asl., but forests with a function other than permanent production forest were excluded from the NFI.

Nowadays in many places, forest areas are fragmented so that they are not covered by the original NFI sampling grid of 20 km x 20 km. Therefore, in 2005 the MoEF started to establish new cluster-plots in the fragmented forest areas by using the sampling grids of 10 km x 10 km or 5 km x 5 km. To increase the spatial coverage of cluster-plots, in 2019, the MoEF has also started to establish new cluster-plots in mangrove forests and mountainous forests with altitudes above 1000 m that were not covered yet by the NFI design (Tatang 2022).

The NFI used a single-stage systematic sampling design (with post-stratification) instead of a two-phase sampling design because remote sensing data (either aerial photographs or satellite images) were challenging to obtain in the 1990s. The design of cluster-plots has been kept unchanged since 1989. Originally, a total of 2,735 plot clusters were enumerated. Each cluster consists of 9 tracks, of which 8 contains 72 temporary plots and one track a 1-ha permanent sample plot (Figure 18). The PSP is further divided into 16 record units (25 m x 25 m on drylands; 12.5 m x 12.5 m on mangroves, Figure 18).

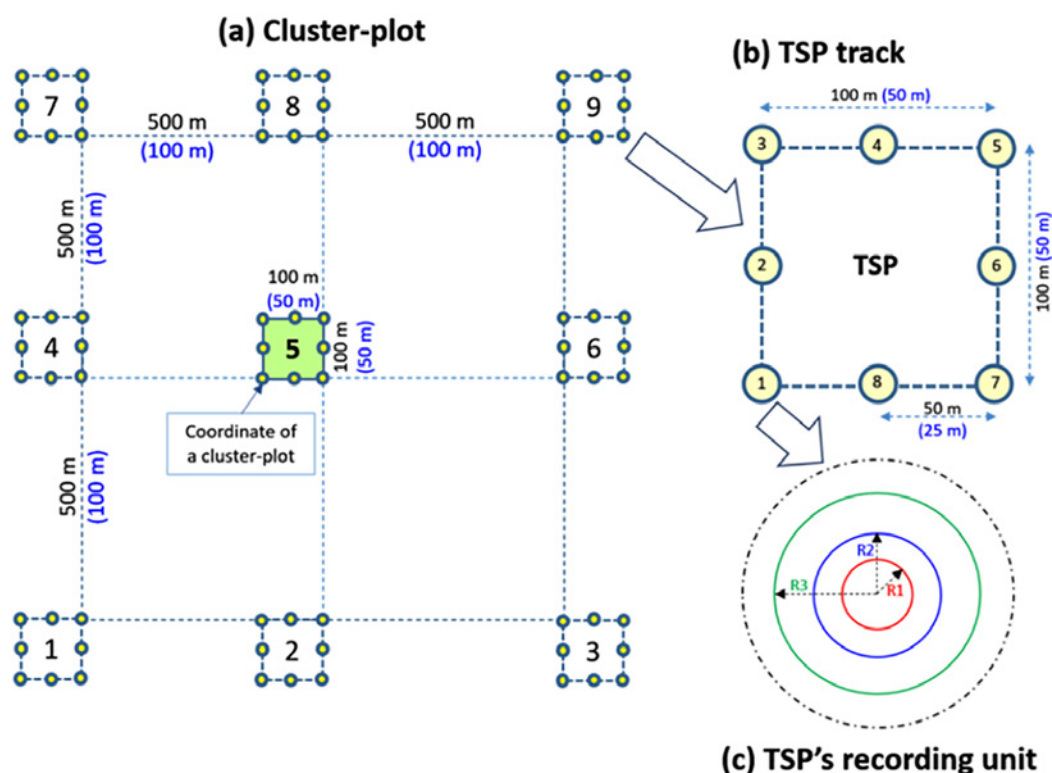


Figure 18. Field sample plots in inventory clusters.

For the NFI design analysis, FAO experts got a subset of old NFI PSP data that was collected between 2000–2020. The dataset contains 16 Excel files: data is split into 7 islands and a one mangrove dataset (for the whole country), and the files are labeled as follows:

- RT45 – data by Record Units (RUs), and
- RT46 – tree and pole data ($\text{DBH} \geq 5 \text{ cm}$).

Because data about the land cover class was mostly missing in RT45 files, land cover class was taken from the MoEF Land Cover maps for each PSP at the time of the assessment. In addition, Forest/Non-Forest class was taken from these maps.

For the analysis, only PSPs falling into the “forest” category were selected, and a total of 5 PSPs (from Sumatra) were dropped due to inconsistencies in the RT46 data. Total of 1,454 PSPs were used in this analysis. The layout of the PSPs in Drylands and Mangroves (where dimensions are in brackets) are shown in Figure 19.

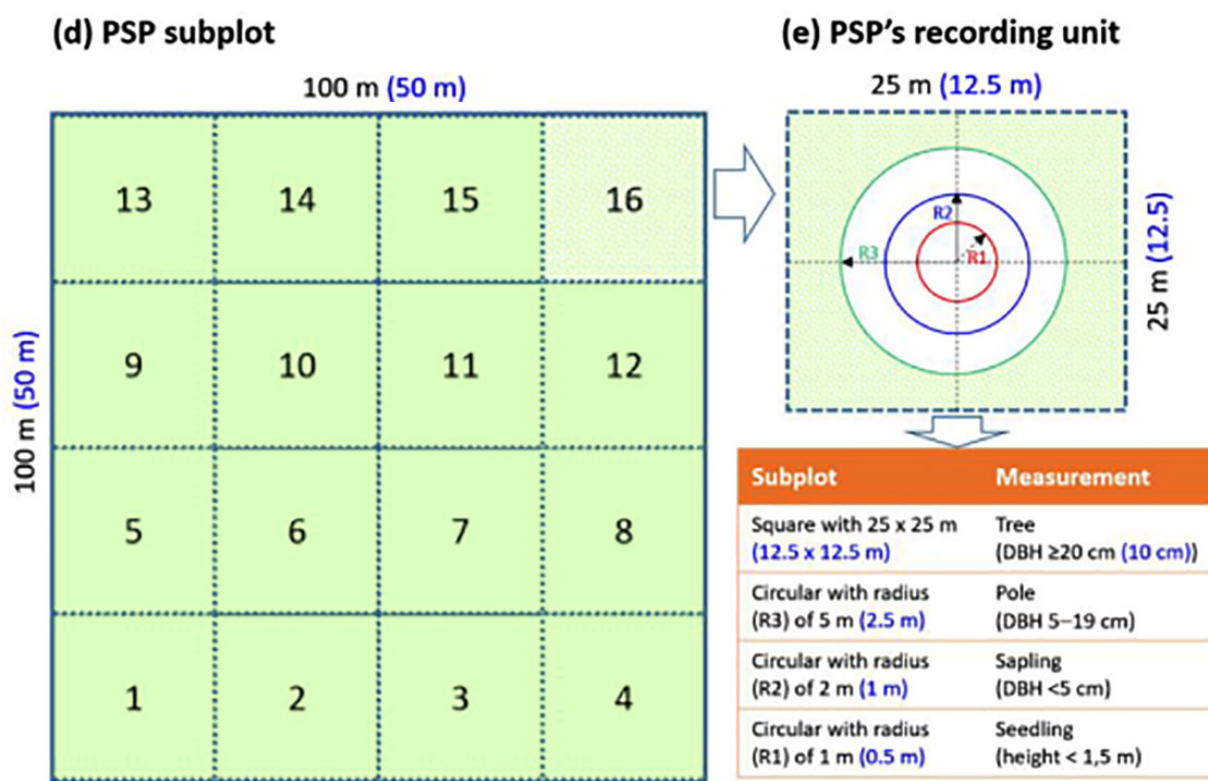


Figure 19. Layout of a Permanent Sample Plot.

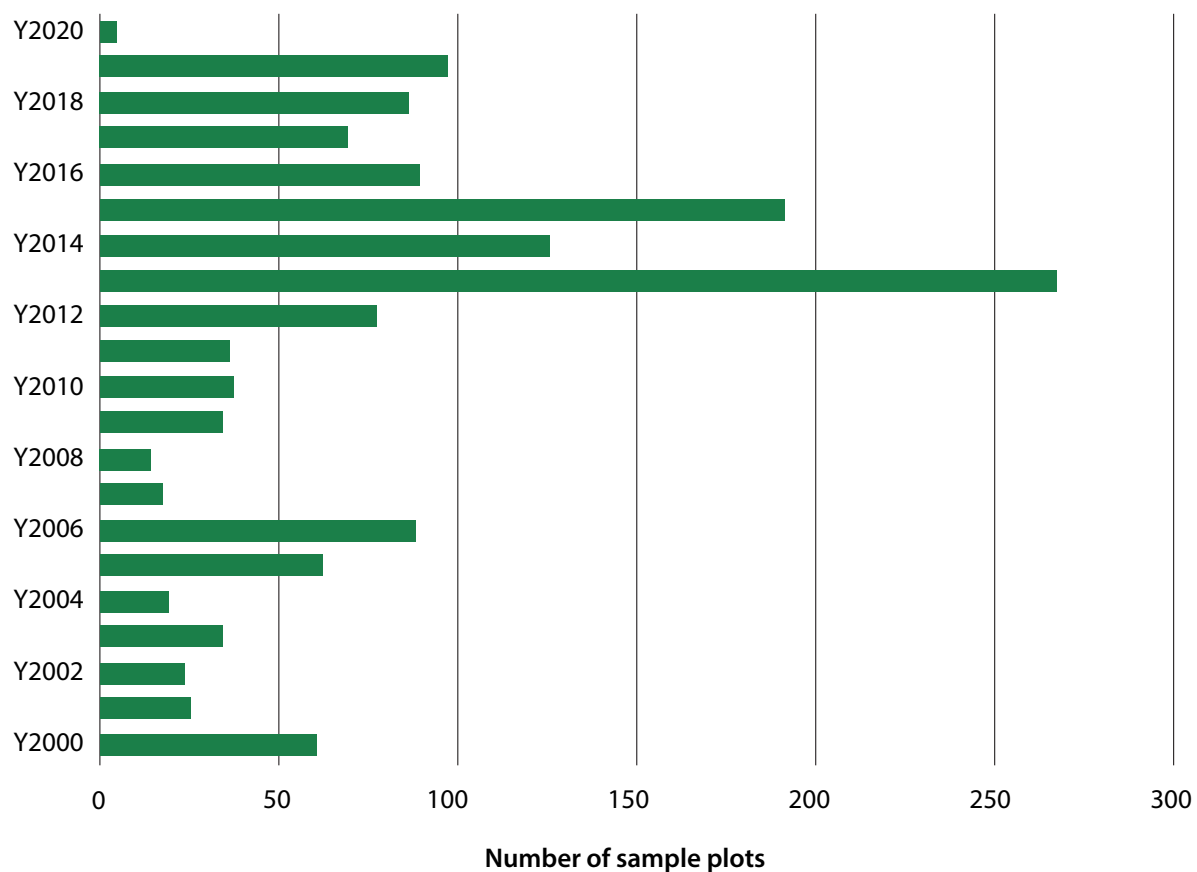


Figure 20. Number of sample plots by year of assessment.

At the first step, tree location maps were used to determine spatial distribution of trees and suitability of “tree maps” for sampling simulations. However, based on visual assessment of mapped trees, trees were determined to be too clustered in the data (Figure 21) and it was decided to use directly the record unit level data.

DBH distributions by different land categories were also plotted and these were compared to distributions that we would get by changing the nested plot dimensions and DBH thresholds of these nests. Hence, it was also possible to examine the average number of trees to be recorded by alternative plot designs.

The following charts visualize the relative frequencies of trees in 2 cm diameter classes, up 60 cm DBH on drylands by islands, and in mangrove forests in the NFI data. The average numbers of recorded trees in a whole PSP and in RUs are presented in the text boxes.

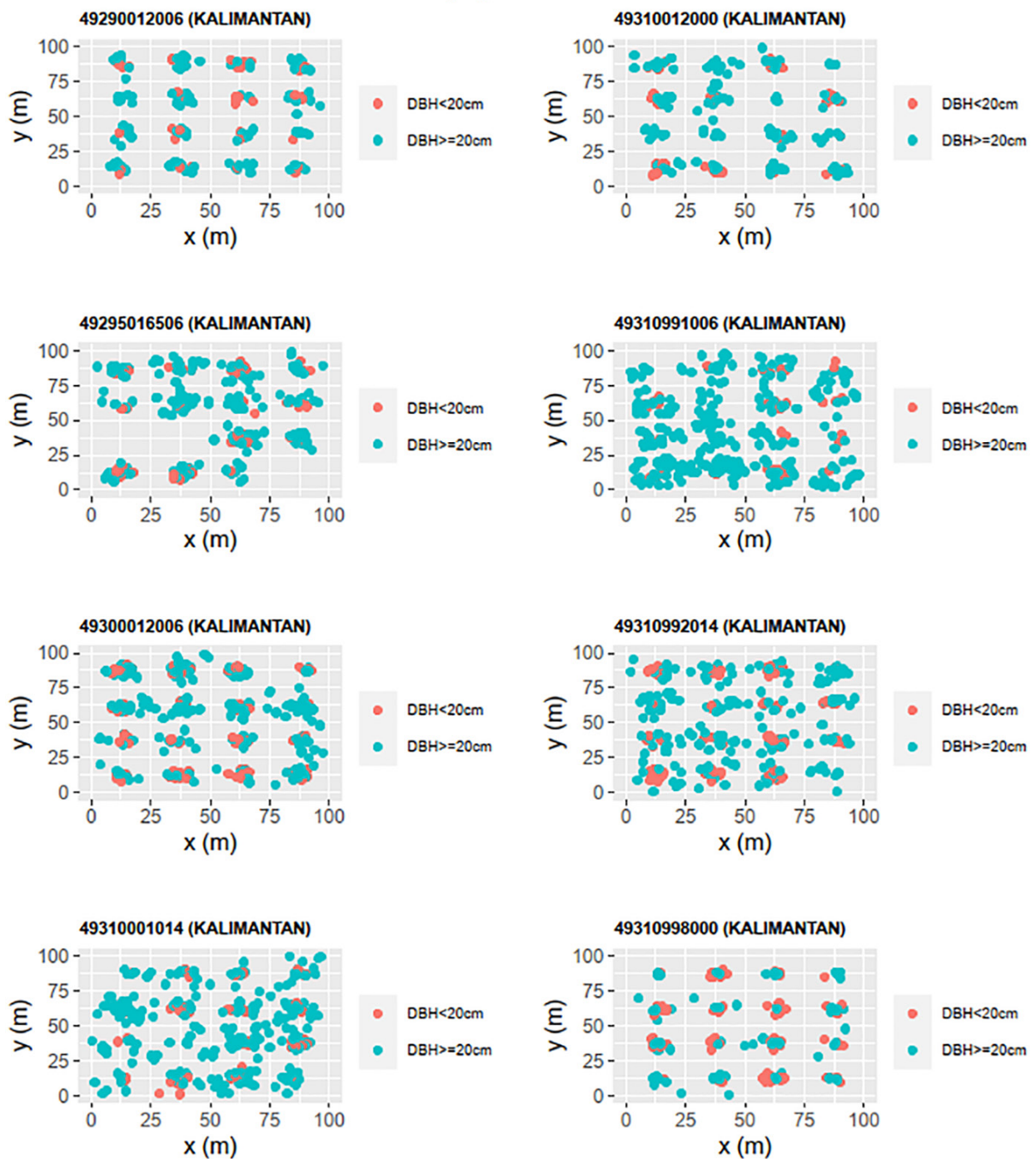


Figure 21. Demonstrated clustering effect in permanent sample plots within a subset of plots in Kalimantan data.

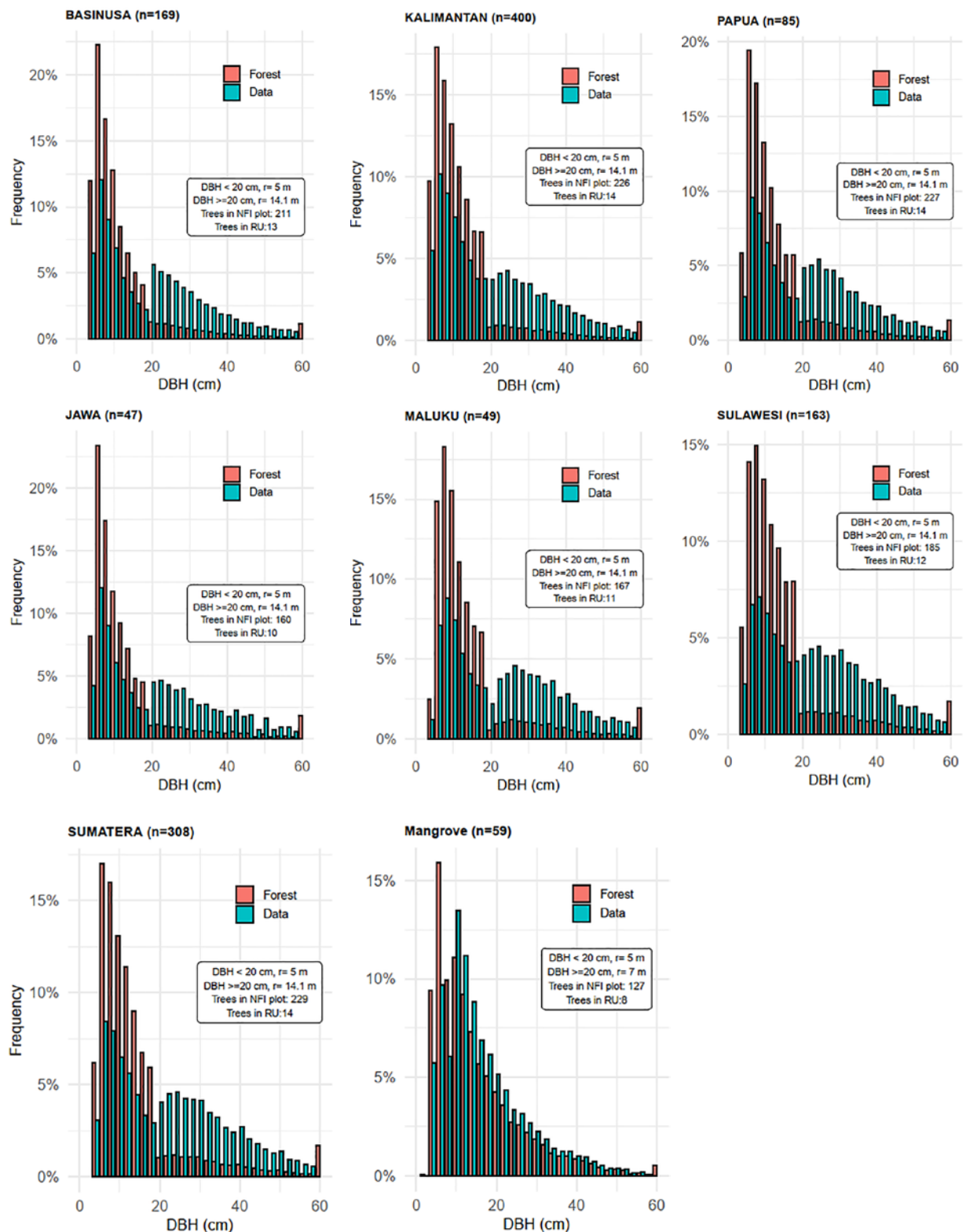


Figure 22. DBH distributions by islands (drylands only) and in mangrove forest. The turquoise bars represent the recorded trees in RUs. Dimensions of RU squares are presented as in equal size circles (text boxes).

The DBH distribution graphs show that small diameter classes are slightly over-represented in the drylands data, but more strongly in the mangrove data. However, because in an even-aged forest the diameter distribution is always a declining curve, so-called reverse J shape, the assumption to cover all DBH classes evenly is not realistic to achieve practice, but it can still be an appropriate criterion when choosing the plot type(s).

In order to improve the plot layout, we examined several alternative options by using the PSP data (Table 11).

Table 11. Examined concentric circular plot designs.

Option	Sites	Nest 1	Nest 2	Nest 3
1	Drylands, Mangrove	DBH<15cm r = 5 m	DBH>15cm r = 15m	
2	Drylands, Mangrove	DBH<25cm r = 5 m	DBH>25cm r = 20m	
3	Drylands, Mangrove	DBH<20cm r = 5 m	DBH>20cm r = 18m	
4	Drylands, Mangrove	DBH<20cm r = 5 m	DBH>20cm r = 20m	
5	Drylands, Mangrove	DBH<20cm r = 5 m	20cm<DBH<40cm r = 15m	DBH>40cm r = 20m
6	Drylands, Mangrove	DBH<20cm r = 5 m	20cm<DBH<40cm r = 15m	DBH>40cm r = 25m
7	Drylands, Mangrove	DBH<15cm r = 5 m	15cm<DBH<25cm r = 10m	DBH>25cm r = 20m
8	Mangrove	DBH<10cm r = 5 m	DBH>10cm r = 8m	
9	Mangrove	DBH<10cm r = 5 m	DBH>10cm r = 10m	
10	Mangrove	DBH<15cm r = 5 m	DBH>15cm r = 10m	
11	Mangrove	DBH<10cm r = 5 m	10cm<DBH<20cm r = 8m	DBH>20cm r = 12m

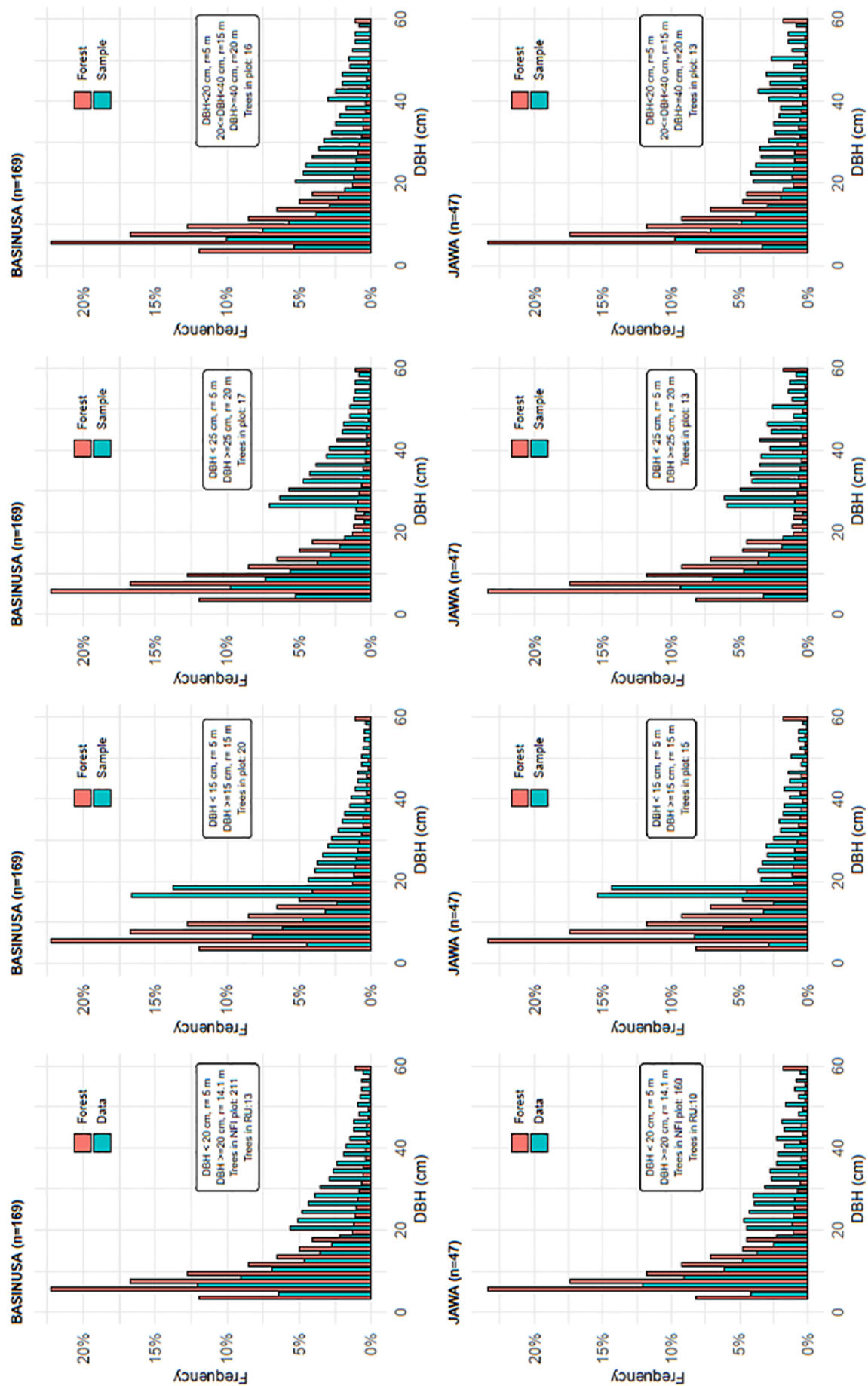
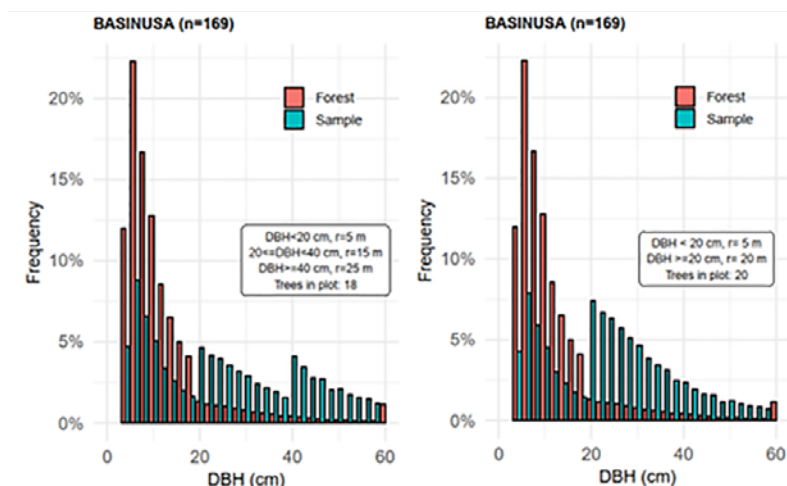


Figure 23. Examples of computed DBH distributions with different plot design options for Bali Nusa and Jawa.

Two best performing options were #4 and #6 for dryland forest, and #9 and #11 for the mangrove forests (Figure 24).

Drylands



Mangroves

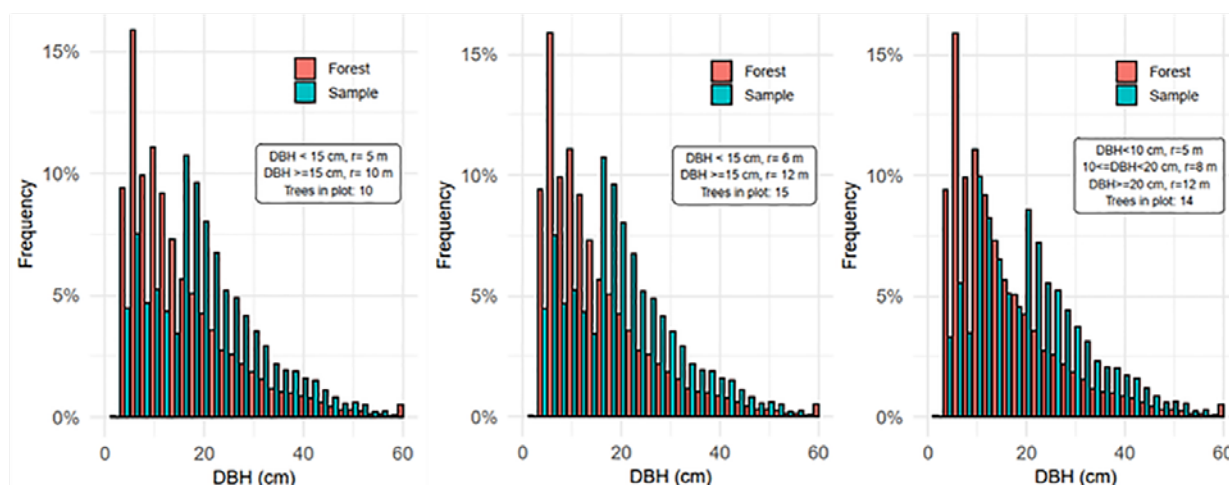


Figure 24. Plot design options for further analysis (drylands and mangroves across all islands).

Plot size and cluster plots

The targeted number of plots depends on variation of the target variable in the population. The primary target variable is the above-ground biomass (AGB), but because basal area is strongly correlated with AGB we used that when examining the variance graphs. Thus, the effect of the plot size to the Coefficient of Variation (CV) of basal area was examined next. First, we generated a new “sampling data” based on RU level data from PSPs. The original PSPs were used as “realization of reality” and selected designs were randomly allocated into these populations. We preselected PSPs that were located in forest at time of the assessment and which contained at least 10 record units. Inside each PSP, we sampled one RU for small trees (i.e. one circular plot with 5 m radius, DBH less than 20 cm, and 1–5 RUs for trees DBH greater or equal to 20 cm for drylands).

Similar sampling method was also applied with mangrove data, where inside each PSP, we sampled one RU for small trees (i.e. one circular plot with 5 m radius and DBH less than 10cm, and 1–5 RUs for trees DBH greater or equal to 10cm).

Original RU information was used to demonstrate alternative dimensions of field plot circles, and cumulative areas of RUs compared to a one circular plot is presented in the next table. When 3 RUs was used, it was referring to a circular sampling area with a radius of 24.4 m, which is close to the planned 25 m maximum radius on drylands. Similarly, if maximum radius for a circular plot in mangrove data was set to 12 m, this corresponds cumulative area of 3 RUs (Figure 25).

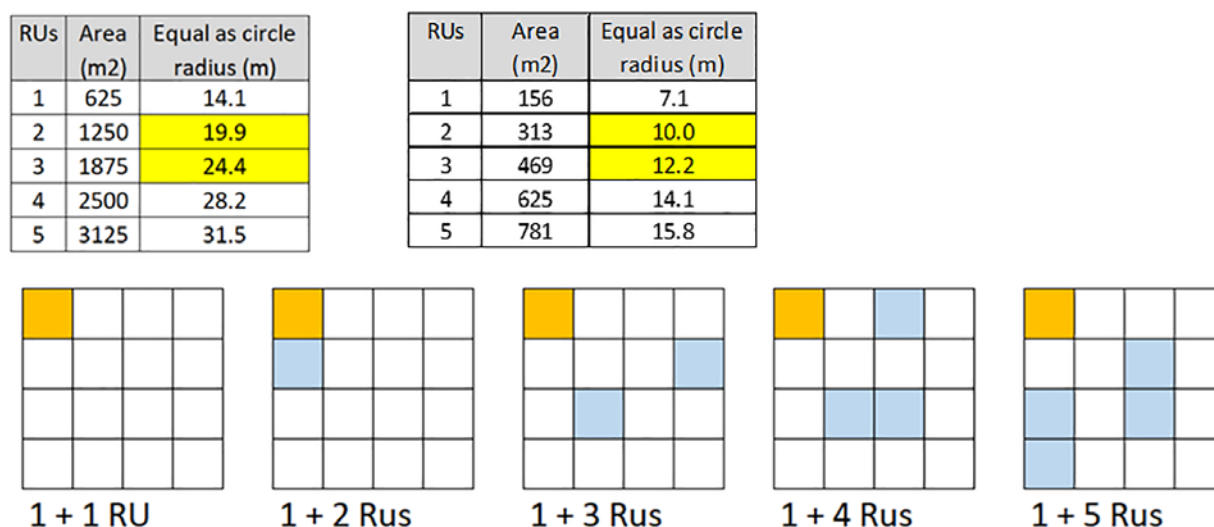


Figure 25. Examined plot design options for drylands and mangrove sites with a plot generated from randomly chosen RUs.

Randomized simulations were applied using a scheme described in Figure 25 so that into each PSP we simulated 10 new “circular plots”, and computed the basal area (per hectare) for each generated “plot”. Next, the CV of the basal area was analyzed with different plot and cluster-plot combinations.

Cluster plots are a special form of fixed area plots that are composed of spatially dispersed subplots. Cluster plots can be integrated with other plot designs. Cluster plots have the advantage that more of the small-scale variability can be captured inside the plot compared to a single plot (Fehrmann *et al.*, 2017).

A cluster-plot in this study was generated by sampling one to two additional “circular plots” inside each PSP (with random sampling without replacement). It was clear that when the sampled area increases (i.e. two or three plots are combined), the variation decreases. Because cluster plots actually form one “super-plot”, we could simulate the effect of increased sample area to variation with the help of the PSP data. The next charts show the CV of the basal area (estimate) with different plot sizes (x-axis) and with 1 plot or 2–3 cluster plots (these are called “subplots” in the charts).

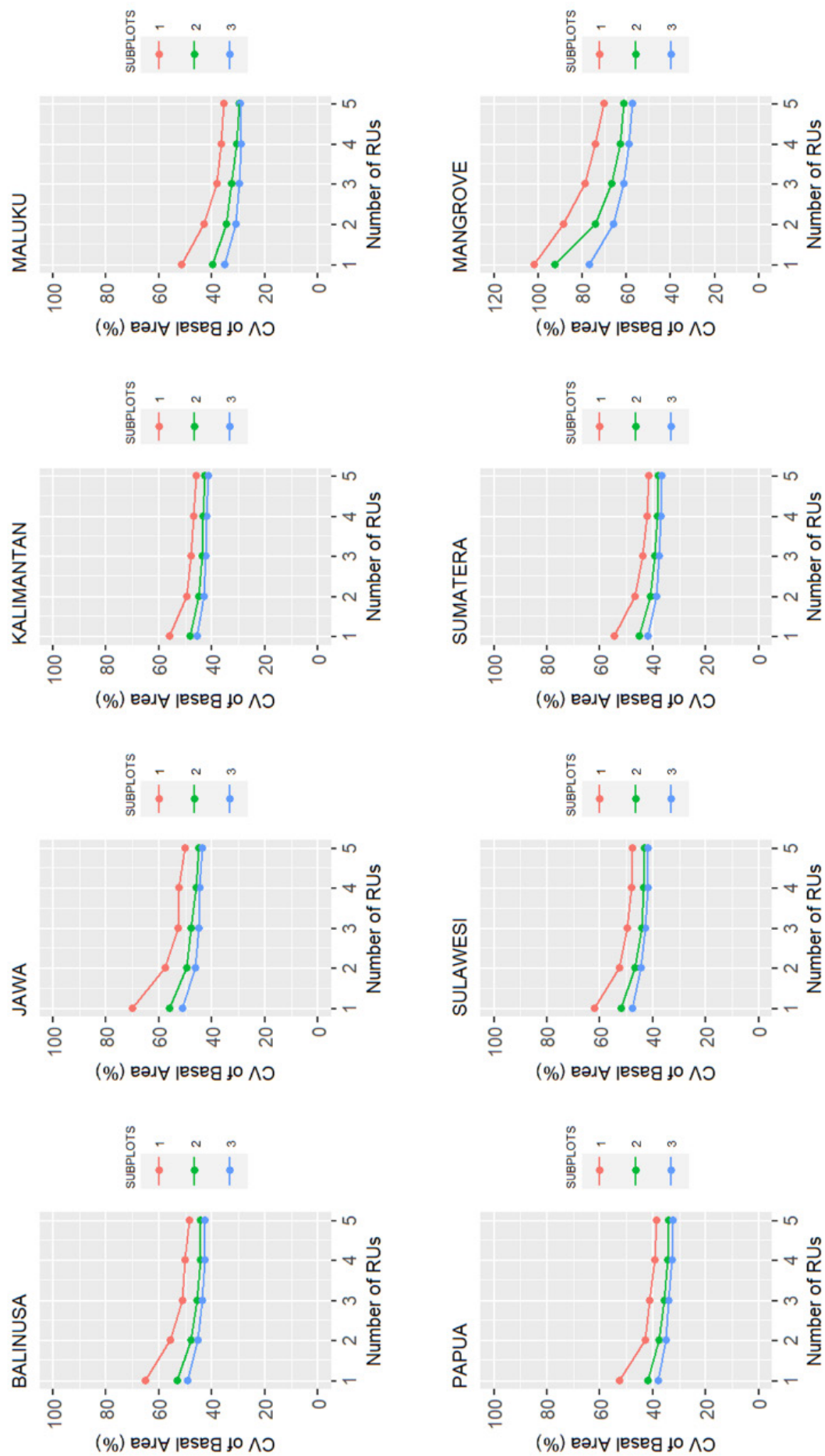


Figure 26. CV of basal area analyzed with different plot and subplot combinations.

In all dryland forest charts above (from Bali Nusa to Sumatra), CV sets down to quite constant level with 3 subplots and sampling area of 3 RUs (which corresponds 24.4 m circular plot area). At mangrove sites, the trend line continues declining, but not strongly when the number of subplots is equal to three.

Another consideration is what can be a maximum plot radius in a tropical rainforest. Based on experts' experiences in working in Kalimantan, even 20 m is a long distance for a circle when using a measuring tape. On the other hand, bigger plots will better capture very big trees, that are relatively rarer than small trees in forests, and which can contain a major share of biomass. Luckily new devices, such as electronic rangefinders and Vertex 4, can calculate horizontal distances and show them for the measurer. Therefore we set two options for the maximum plot radius: 1) 20 m that approx. matches with cumulative area of 2 RUs, and 2) 25 m that approximately matches with cumulative area of 3 RUs. We do not recommend applying longer distances for plot radii.

By choosing the plot design option #6 for dryland sites, the cumulative area in a plot is 0.589 ha which can be regarded sufficient to tropical forest to capture species composition and diameter distribution of trees⁹. This is also in line with the results of Nanjaya *et al.* (2020) who presented optimal plot sizes of 0.40 ha for Primary Dryland Forest (PDF), Secondary Dryland Forest (SDF), and Secondary Peatswamp Forest (SPF), and 0.25 ha for Primary Peatswamp Forest (PPF). Tree measurements should contain both live and dead standing trees and palms.

Number of samples

A smaller number of plots is often acceptable in locations where there is relatively low variability in stand density (in terms of biomass in this paper) and composition; more plots are needed in stands with high variability. The minimum required number samples for different plot combinations was analyzed with a tailored FAO tool¹⁰ both with the bootstrap method and Cochran's formula (1977):

$$n_{i+1} = \left(\frac{t_{n_i-1, \alpha/2} CV}{E} \right)^2$$

Where

n = number of samples

CV = Coefficient of Variation

E = accuracy target

With the help of bootstrap algorithm or Cochran's formula we can set the targeted sampling error (as percent) and get out the number of samples and confidence intervals for the estimates of the mean, in this case for above-ground biomass in forest by islands. The FAO tool also creates two plots that provide a view into the increase in the accuracy of estimates of the mean as sampling intensity is increased. Four accuracy targets (in terms of sampling error for the mean biomass estimate) were set in, as follows: 20%, 15%, 10% and 5%.

⁹ This agrees with studies that have reported that a plot size of approximately 0.35–0.5 ha is necessary for sampling tropical forest estimated aboveground biomass and other structural characteristics (e.g. Brown *et al.*, 1995; Clark & Clark, 2000).

¹⁰ https://openforis-shiny.shinyapps.io/bootstrap_r/

First, we computed the estimated number of samples using full 1-ha PSPs (Table 12). These estimates were compared with the estimates for smaller simulated plots. As seen in this table, when we set a 5% accuracy target, we could not get estimates by using the bootstrap method due to high variation or lack of data.

Table 12. Original subset of NFI data: Estimated number of samples on drylands by islands.

Bootstrapping				Cochran			
20%	15%	10%	5%	20%	15%	10%	5%
28	51	111		30	49	109	435
28	47	96		27	48	107	426
28	46	97	377	24	42	94	376
15	24	47		12	21	48	189
17	28	60		15	27	59	235
29	49	109		28	48	107	427
24	39	86	312	21	37	83	332
169	284	606		157	272	607	2.420

Table 13 shows the estimated number of samples for the main plot design option, with two analysis methods.

Table 13. Estimated number of samples on drylands by islands with two methods (25 m radius plot and 3 cluster plots design).

ISLAND	Bootstrapping NG				Cochran			
	20%	15%	10%	5%	20%	15%	10%	5%
BaliNusa	33	55	119	470	30	53	119	476
Jawa	42	70	152	516	37	66	148	590
Kalimantan	29	48	105	410	26	47	105	413
Maluku	18	28	59	221	14	25	55	219
Papua	19	32	66	255	16	29	65	257
Sulawesi	33	56	124	479	31	55	121	486
Sumatera	28	45	98	387	24	43	97	386
Total	202	334	723	2738	178	318	710	2827

When comparing the estimates obtained via the bootstrap method and Cochran's formula, there is not much difference in numbers. In the main section of this document, only the results based on Cochran's formula are presented because the estimates of these two methods are very close to each other, and current data for bootstrapping was not sufficient to produce all estimates at the 5% accuracy level. The estimated number of samples and cumulative area to be sampled with different plot design options are presented in Table 14.

Table 14. Estimated number of samples and cumulative area to be sampled by islands with different plot layouts and cluster plot designs.

Number of samples						Area (ha)			
ISLAND	Legend	20%	15%	10%	5%	20%	15%	10%	5%
BaliNusa	r = 20m (1 subplot)	56	98	220	861	7	12	28	108
BaliNusa	r = 25m (1 subplot)	39	70	156	614	8	14	31	121
BaliNusa	r = 20m (3 subplot)	34	59	133	520	13	22	50	196
BaliNusa	r = 25m (3 subplot)	30	53	119	476	18	31	70	280
Jawa	r = 20m (1 subplot)	55	97	217	867	7	12	27	109
Jawa	r = 25m (1 subplot)	53	95	209	836	10	19	41	164
Jawa	r = 20m (3 subplot)	41	73	161	640	15	28	61	241
Jawa	r = 25m (3 subplot)	37	66	148	590	22	39	87	348
Kalimantan	r = 20m (1 subplot)	37	65	146	577	5	8	18	73
Kalimantan	r = 25m (1 subplot)	34	61	135	537	7	12	27	105
Kalimantan	r = 20m (3 subplot)	27	48	108	431	10	18	41	162
Kalimantan	r = 25m (3 subplot)	26	47	105	413	15	28	62	243
Maluku	r = 20m (1 subplot)	36	64	144	573	5	8	18	72
Maluku	r = 25m (1 subplot)	35	62	138	552	7	12	27	108
Maluku	r = 20m (3 subplot)	19	33	72	286	7	12	27	108
Maluku	r = 25m (3 subplot)	14	25	55	219	8	15	32	129
Papua	r = 20m (1 subplot)	28	48	108	431	4	6	14	54
Papua	r = 25m (1 subPlot)	25	45	100	398	5	9	20	78
Papua	r = 20m (3 subplot)	18	32	71	287	7	12	27	108
Papua	r = 25m (3 subplot)	16	29	65	257	9	17	38	151
Sulawesi	r = 20m (1 subolotl)	53	97	212	856	7	12	27	108
Sulawesi	r = 25m (1 subplot)	42	74	165	658	8	15	32	129
Sulawesi	r = 20m (3 subplot)	32	56	123	497	12	21	46	187
Sulawesi	r = 25m (3 subplot)	31	55	121	486	18	32	71	286
Sumatera	r = 20m (1 subplot)	38	67	150	602	5	8	19	76
Sumatera	r = 25m (1 subplot)	33	59	132	525	6	12	26	103
Sumatera	r = 20m (3 subplot)	27	46	105	411	10	17	40	155
Sumatera	r = 25m (3 subplot)	24	43	97	386	14	25	57	227
Mangrove	r = 10m (1 subplot)	98	174	390	1560	3	5	12	49
Mangrove	r = 12m (1 subplot)	69	121	275	1098	3	5	12	50
Mangrove	r = 10m (3 subplot)	57	100	222	888	5	9	21	84
Mangrove	r = 12m (3 subplot)	48	85	192	762	7	12	26	103

The proposed cluster plot design for drylands is 25 m radius plot with 3 subplots, and for mangroves a 12 m radius plot with 3 subplots. When setting the accuracy target for the mean biomass estimate on drylands to 5%, and for mangroves to 10%, we can get the estimates as presented in Table 15.

Table 15. Estimated number of samples by islands with preferred cluster plot designs.

ISLAND	Plot design	Accuracy	n
BaliNusa	r = 25m (3 subplot)	5%	476
Jawa	r = 25m (3 subplot)	5%	590
Kalimantan	r = 25m (3 subplot)	5%	413
Maluku	r = 25m (3 subplot)	5%	219
Papua	r = 25m (3 subplot)	5%	257
Sulawesi	r = 25m (3 subplot)	5%	486
Sumatera	r = 25m (3 subplot)	5%	386
Mangrove	r = 12m (3 subplot)	10%	192
Total			3.019

2.2 Description of forest structure and scale in forests

2.2.1 Structure of forest areas

In general, on the national level, both correlation and autocorrelation functions have been applied to examine structures of large forest areas, and structural pattern of volume in trees. For example in Scandinavian forests, there is a slightly increasing autocorrelation almost up to a 200 m distance. This information has been used to estimate optimal distance between inventory tracts, overall shape of the tract, and distance between sample plots within a tract. In the case of the Swedish NFI, in order to obtain a sufficient sampling set-up for the entire country, the land is divided into five regions with marginally diverse correlation functions for different variables. The final sampling design is based on their spatial characteristics and other practical considerations.

In this study, we used the following two datasets in the spatial analysis for designing the NFI sampling layout:

1. Global canopy height for the year 2020 and at the 10 m ground sampling distance, and
2. GEDI 2020–22, at the 25 m ground sampling distance.

In general, the existing materials (Land cover map; ETH forest cover and canopy height maps, and optionally Landsat-based data; GEDI sample lines) are good-quality data sources (Figure 27). These can be used to compile a proper stratification and cluster sampling approach. Existing field plot data can be used to design suitable cluster and field plot design.

Forest structures were analyzed separately for each Island at the following levels and distances: plot level (<100m) and cluster level (50–500m). In these data, the relative spatial distribution of forests and trees varies, because of spatial variation in land use patterns, soil fertility, hydrology, and size distribution of trees.

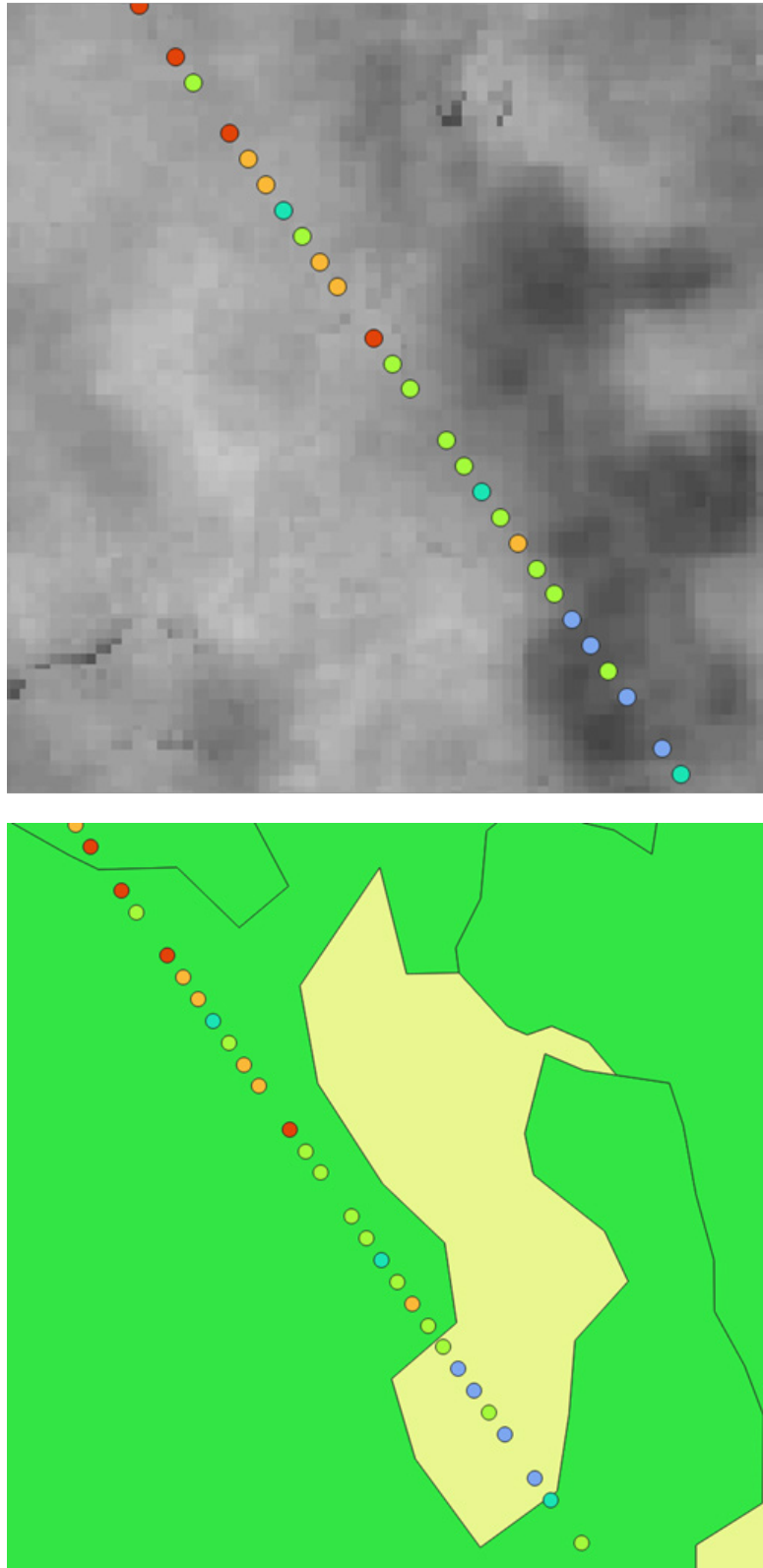


Figure 27. GEDI data (dots) overlaid with canopy height data (left). Variation within forest areas was used in analysis. Also, GEDI AGB was analyzed with land use map categories (right).

The correlograms or variograms have also been used to estimate the standard error as regards large areas and describe spatial autocorrelation (Figure 28). **Autocorrelation is indicating similarity between locations. Distance between useful plots can be done using autocorrelation. If we use more dense network plots, we are practically measuring the same information. Thus, when we have very small coverage with a good-quality image, the information content is limited if similar forest areas are covered inside the scene.** The standard errors of systematic cluster sampling can be estimated using model-based estimators, which utilize the parameters of correlation functions. In that way, the information about spatial dependency can be utilized in the error estimation and spatial structure of forests is taken into account.

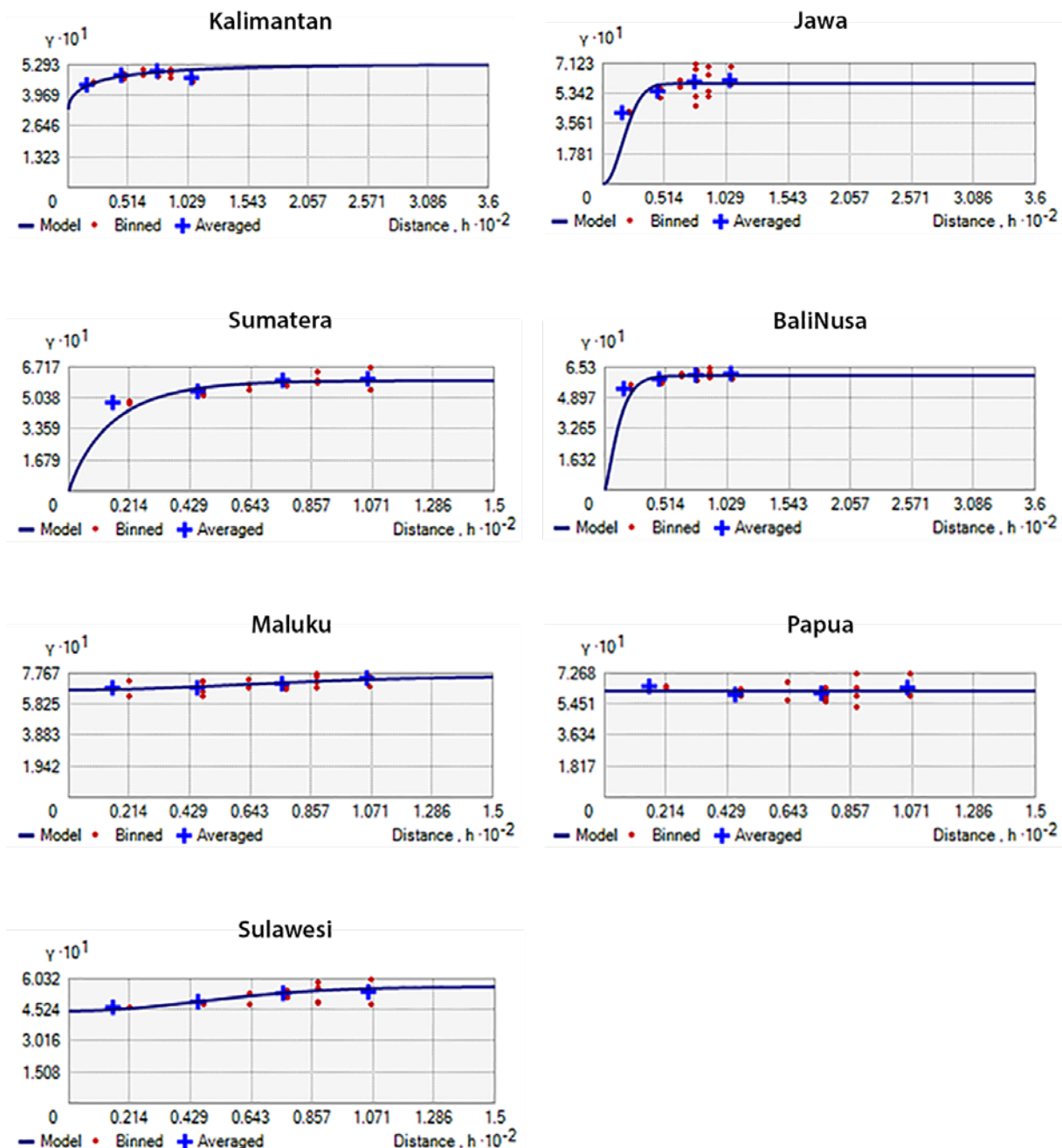


Figure 28. Autocorrelation inside PSPs in Islands.

Autocorrelation inside field plots was checked to review short-distance variation. In Kalimantan, Jawa, Sumatera, and Bali Nusa data, autocorrelation figures after 50 m distance are very random. Maluku, Papua, and Sulawesi data did not have enough observations for this analysis or variation was too high. So, no clear autocorrelations were observed. When the intracluster correlation of large trees (DBH ≥ 40 cm) above-ground biomass was calculated for PSP plot record units, and the following ICC results were obtained (Table 16). The reason for examining big trees is that these trees are important for tropical forest biomass storage, and that density of large trees can explain the majority of the variation in pan-tropical AGB (Slik et al., 2013).

Table 16. Intracluster correlation of large trees (DBH ≥ 40 cm) above-ground biomass calculated for PSP record units.

Kalimantan	0.32
Sumatera	0.60
Jawa	0.20
BaliNusa	0.22
Sulawesi	0.28
Maluku	0.14
Papua	0.20

The conclusion here is that these are very low correlation figures and there is no reason to apply long distance between the sample plots (just 50 m is enough), and applying larger plots can be justified. When cluster level distances (100–500 m) were analyzed, variograms and covariance figures were estimated separately for each island (Figure 29 and 30). Some kind of correlation was observed up to 200–350 m, depending on the island.

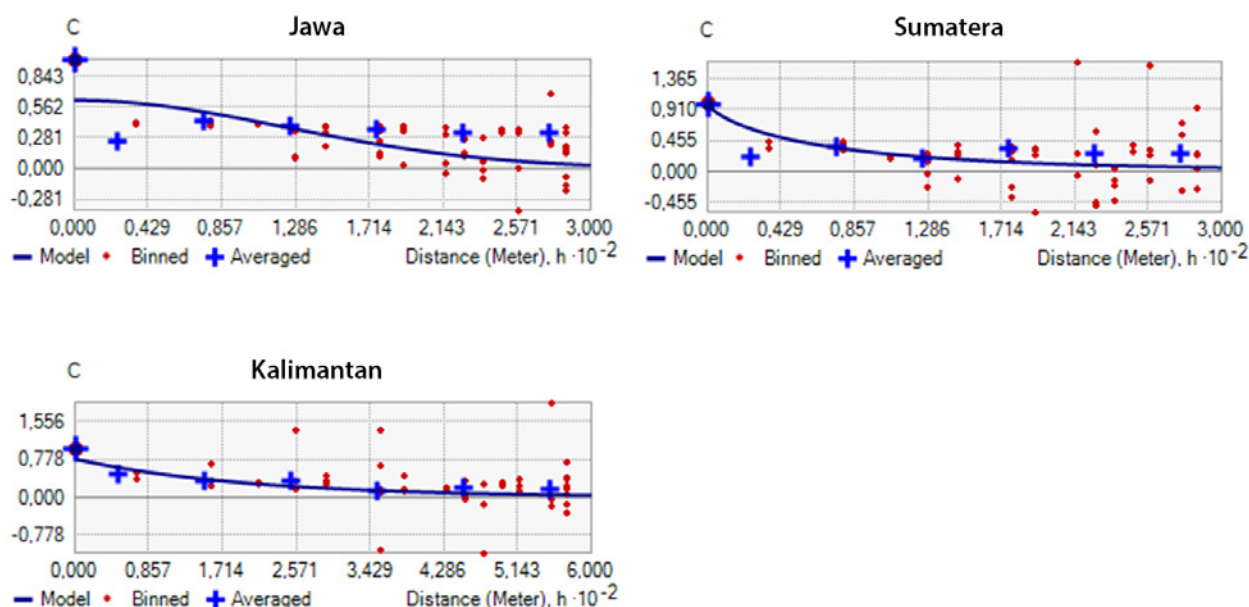


Figure 29. Covariance (correlation) of GEDI data (AGB) in inside forest by islands.

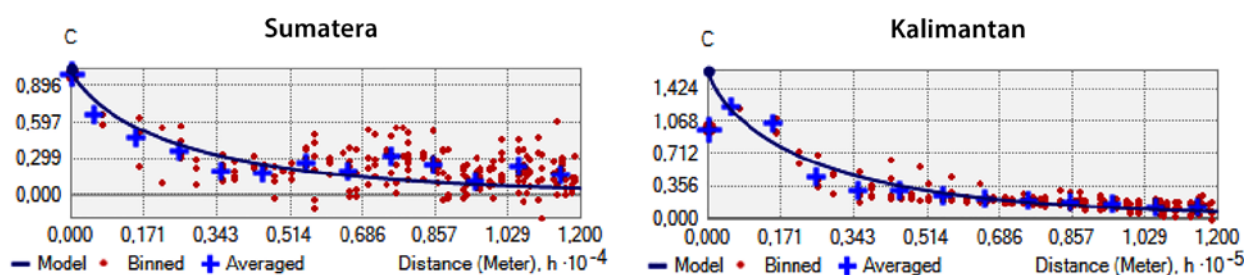


Figure 30. Covariance (correlation) based on ETH canopy height values.

Jawa and Bali Nusa areas are different areas compared to other Islands, but differences are small if we focus only inside forest variation. Land use is very correlated when forest area is small.

2.2.2 Forest types and canopy cover in Land Use Maps

In forests, there are many variables that are spatially sparse and scattered. This is often the case when one is assessing coarse woody debris in managed forests, or surveying threatened species. The spatial description of sparse populations can also be problematic. On the landscape level, information about spatial distribution of different key habitats and areas with a high ecological value has also been used to assess the probability of existing rare species. Field data about indicator species and remote sensing data about landscape features are valuable *a priori* information for estimating the presence/absence probability and for stratifying areas of interest.

Table 17. GEDI AGB in various Land use/Land cover types in Kalimantan.

Variable Code	GEDI, AGB, Kalimantan				Cover, %	
	Penggunaan/penutupanlahan	Landuse/landcover	Average	StdDev	Average	Std Dev
2001	Hutan Lahan Kering Primer	Primary dryland forest	221.9	132.1	92	7.6
2002	Hutan Lahan Kering Sekunder	Secondary dryland forest	199.6	129.3	90.6	10.7
2004	Hutan Mangrove Primer	Primary mangrove forest	87.5	90.9	89.3	16.1
2005	Hutan Rawa Primer	Primary swamp forest	128	72.7	92.4	8.6
2006	Hutan Tanaman	Plantation forest	89.2	68.9	77.5	25.6
2007	Semak/ Belukar	Dry shrub	107.2	81.8	81.6	21.7
2010	Perkebunan	Estate crop	62.8	62	59.4	35.4
2012	Pemukiman	Settlement areas	54	37.5	32.4	35.7
2014	Tanah Terbuka	Bare ground	49.5	50.1	73.2	31.8
3000	Savana/ Padang Rumput	Savanna and grasses				
5001	Tubuh Air	Open water	97	100.6	56.9	39.7
20041	Hutan Mangrove Sekunder	Secondary mangrove forest	80.6	50.4	85.7	18.7
20051	Hutan Rawa Sekunder	Secondary swam forest	121.1	75.2	92.9	7.7
20071	Belukar Rawa	Wet shrub	60.6	62	77.9	23.9
20091	Pertanian Lahan Kering	Pure agriculture	58.5	50.2	64.6	32.8
20092	Pertanian Lahan Kering Bercam	Mixed dry agriculture	95.3	70.7	77.3	24.7
20093	Sawah	Paddy field	45.8	83.2	46.5	34.6
20094	Tambak	Fish pond/aquaculture	45.3	31.1	50.6	36.6
20121	Bandara/ Pelabuhan	Port and harbor				
20122	Transmigrasi	Transmigration areas	67.2	94.7	40.7	38.5
20141	Pertambangan	Mining areas	60.3	93	42.9	39.9
50011	Rawa	Open swamp	45.2	41.1	64.1	31.7

Structure by forest types was studied using GEDI AGB biomass estimates. Land use category specific biomass and forest cover estimates were calculated for Kalimantan (Table 17) and Jawa (Table 18). In Kalimantan, dry land forest categories (codes: 2001, 2002, 2005, 20051) typically AGB over 100 tons/ha. Mangrove forest categories had slightly lower content. It can be observed that Dry shrub land also had a remarkable amount of biomass. This is caused by difficulties in delineating forest from shrub lands. A similar situation was observed in Jawa where there is a remarkable amount of canopy cover in all land uses (Table 18).

Table 18. GEDI AGB in various Land use/Land cover types in Jawa.

Variable		GEDI, AGB, Kalimantan			Cover, %	
Code	Penggunaan/penutupanlahan	Landuse/landcover	Average	StdDev	Average	Std Dev
2001	Hutan Lahan Kering Primer	Primary dryland forest	200.1	116.7	88.9	6.3
2002	Hutan Lahan Kering Sekunder	Secondary dryland forest	166.7	112.5	80.7	19.2
2004	Hutan Mangrove Primer	Primary mangrove forest	66.5	75.1	85	19.4
2005	Hutan Rawa Primer	Primary swamp forest				
2006	Hutan Tanaman	Plantation forest	104.9	82.8	50.6	34.7
2007	Semak/ Belukar	Dry shrub	117.9	105.3	55.2	34.1
2010	Perkebunan	Estate crop	86.5	62.8	56.7	32.9
2012	Pemukiman	Settlement areas	60.7	53.1	25.8	29
2014	Tanah Terbuka	Bare ground	78.9	60.9	39.7	32.3
3000	Savana/ Padang Rumput	Savanna and grasses	83.8	57.2	45.2	32.3
5001	Tubuh Air	Open water	87	224.9	22.9	31.6
20041	Hutan Mangrove Sekunder	Secondary mangrove forest	40.7	62.6	57.3	35.4
20051	Hutan Rawa Sekunder	Secondary swam forest				
20071	Belukar Rawa	Wet shrub				
20091	Pertanian Lahan Kering	Pure agriculture	77	67.6	39.2	33.1
20092	Pertanian Lahan Kering Bercampur	Mixed dry agriculture	95.6	60.5	52.2	33.1
20093	Sawah	Paddy field	61.1	60.6	28.4	30.7
20094	Tambak	Fish pond/aquaculture	21.6	16.9	8.2	19.8
20121	Bandara/ Pelabuhan	Port and harbor	22.9	14	17	20.7
20122	Transmigrasi	Transmigration areas				
20141	Pertambangan	Mining areas	52	50.5	22.2	28.9
50011	Rawa	open swamp				

2.3 Comparison cluster sampling alternatives using ICC

Cluster sampling is widely used in large area forest inventories, such as in NFIs. The major reason does not lie in the statistical performance but in practical aspects: once a field team has reached a field plot location (which in some cases may take many hours or even some days), then one wishes to collect as much information as possible at this particular sample point. The major point in cluster sampling is that for each random selection not only one single sampling element is selected but a set (cluster) of sampling elements; thus, a cluster consists of a group of observation units, which together form a sampling unit.

When the tree population is spatially having large variation, we prefer independent observation within one day with a feasible distance (100–500 m). The initial selection itself can be done according to any sampling design (simple random, systematic, stratified etc.). The spatial configuration of cluster plots (in terms of number of subplots and their distance) is a matter of optimization based on the Intraclass Correlation Coefficient (ICC) and practical considerations on costs. The concept for calculating cluster sampling-based sample size is based on ICC is following:

- Step 1: Calculate Simple Random Sample Size (SRS).
- Step 2: Calculate ICC – Calculated r_l , Number of Clusters
- Step 3: Calculate Design Effect, $DE=1+(m-1)r_l$
- Step 4: Multiply SRS with DE
- Step 5: Calculate the minimum clusters required

It was obvious that 3 subplots are applicable sampling units for the field work. The major question was, if subplots should be measured from short distances (like large plots) or long distances (like track or cluster). When intraclass correlations were calculated in different Islands, it was clear that short distances are producing lower number plots. When distances 250–350 meters with subplots are applied, the total amount of subplots would be 9,529. When short distance (50–100 m) is applied, total amount of subplots is 8,481 (Table 19).

Table 19. Alternative plot layouts and Intraclass Correlation Coefficient analyzed with different cluster designs.

Concept		N. Plots required			Plotarea	
ISLAND	Legend	10%	5%	DE*n	without DE	with DE
BaliNusa	r = 25m (1 subplot)	156	614	1109	614	1109
BaliNusa	r = 25m (3 subplot)	119	476		1428	1428
Jawa	r = 25m (1 subplot)	209	836	1583	836	1583
Jawa	r = 25m (3 subplot)	148	590		1770	1770
Kalimantan	r = 25m (1 subplot)	135	537	1452	537	1452
Kalimantan	r = 25m (3 subplot)	105	413		1239	1239
Maluku	r = 25m (1 subplot)	138	552	1289	552	1289
Maluku	r = 25m (3 subplot)	55	219		657	657
Papua	r = 25m (1 subplot)	100	398	1024	398	1024
Papua	r = 25m (3 subplot)	65	257		771	771
Sulawesi	r = 25m (1 subplot)	165	658	1807	658	1807
Sulawesi	r = 25m (3 subplot)	121	486		1458	1458
Sumatera	r = 25m (1 subplot)	132	525	1265	525	1265
Sumatera	r = 25m (3 subplot)	97	386		1158	1158
					Large Plot	8481
					Small	4120
						9529

This means that cluster sampling carries a larger error variance than simple random sampling with the same number of subplots. The planner, however, strives to keep the intracluster correlation as small as possible in order not to lose too much precision.

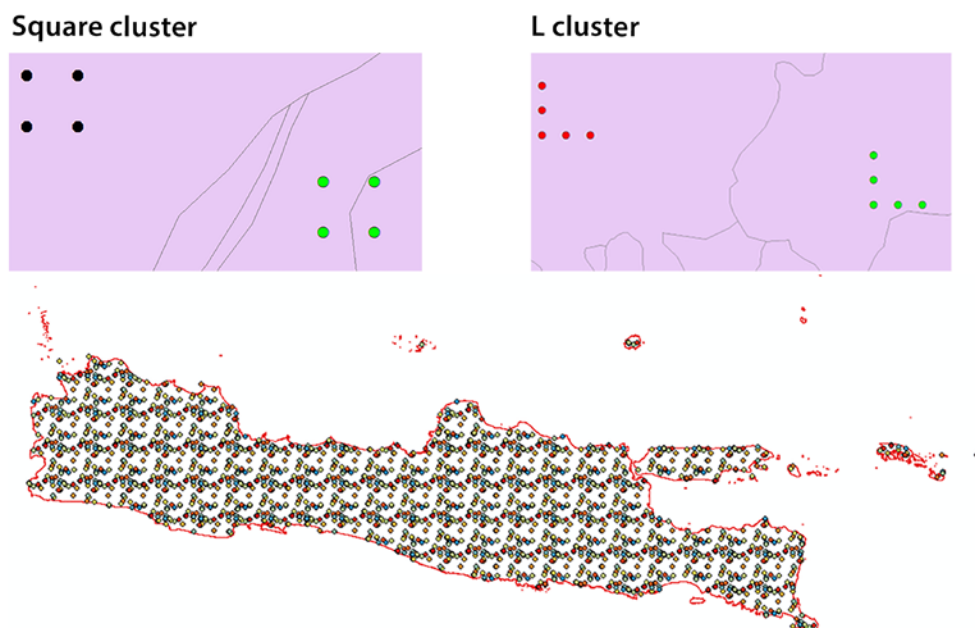


Figure 31. Regular pattern grid was generated to study intracluster correlation with various cluster designs.

A spatial simulator was used to generate alternative cluster designs over a GIS raster layer (i.e., ETH Canopy Height estimates). Simulator was a Python code using *gdal*-functions. It is able to generate the required number of systematic samples over the target area with user defined cluster variation and shape (Figure 32).

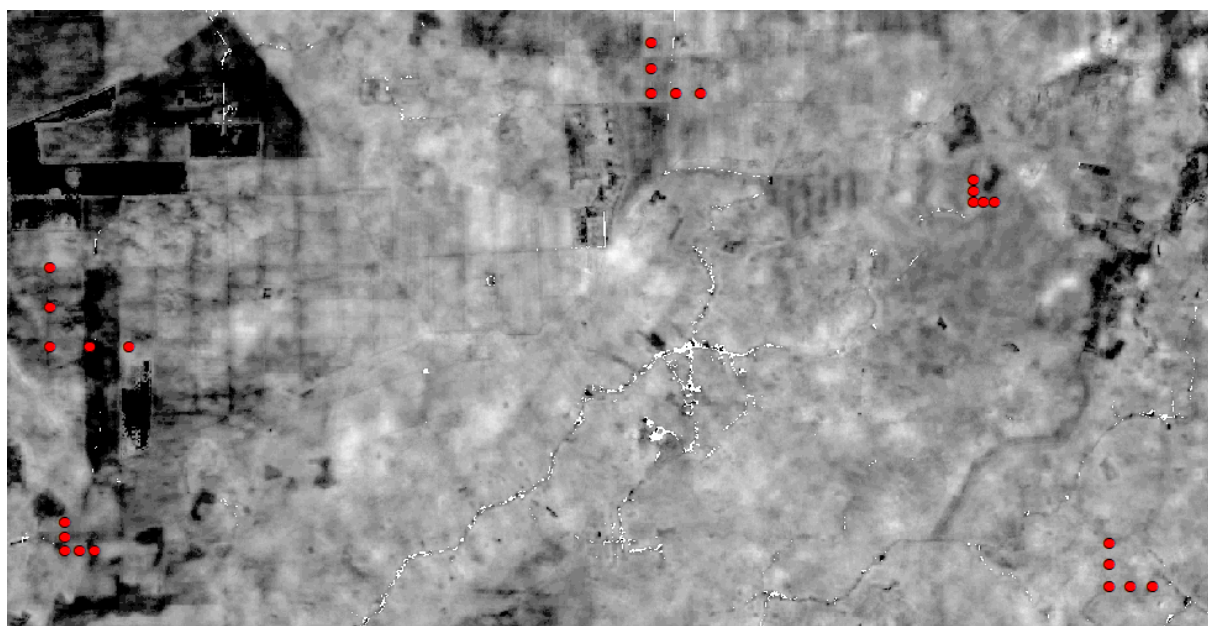


Figure 32. Regular pattern grids (with varying size of clusters) were generated to study intracluster correlation using tree height data, Distance between plots was varying between 50–500 m.

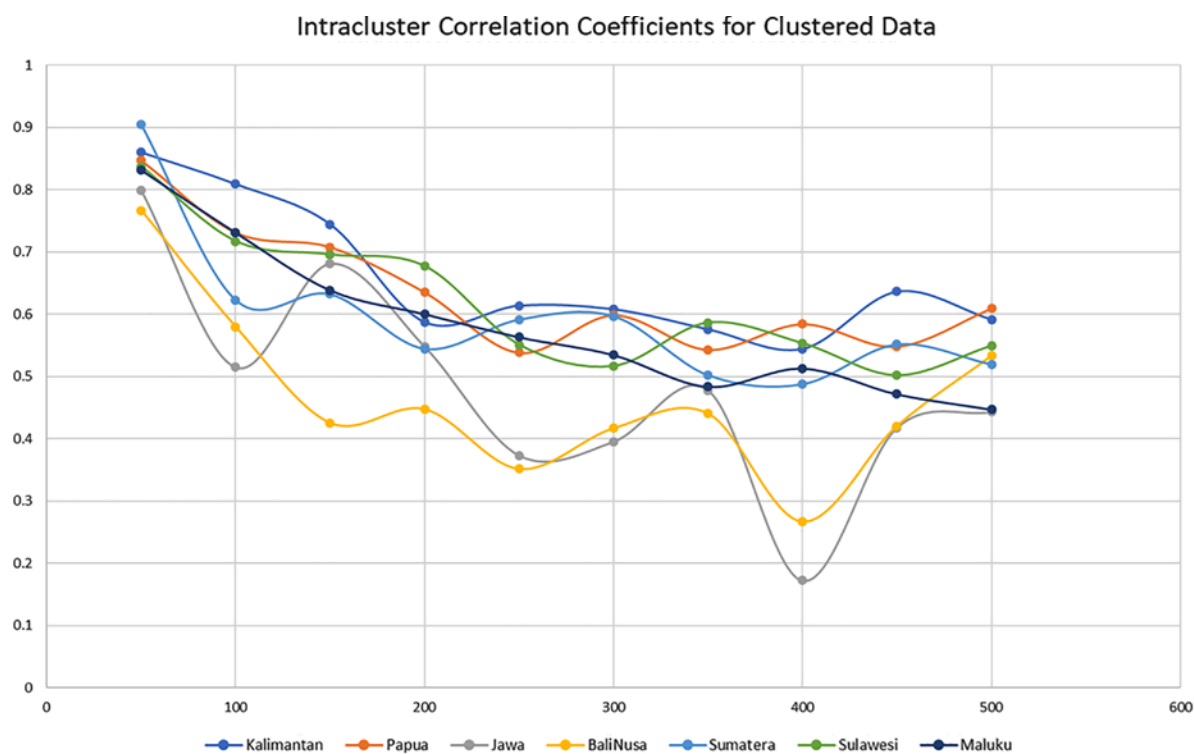


Figure 33. Intracluster correlation curves calculated for islands. X-axis is distance between plots and Y-axis is correlation between plots. A plot is equal to a canopy height observation from the ETH data.

ICC was calculated with the R-function “*clus.rho*”. Island specific ICCs are visualized in Figure 33. Bali Nusa areas had a decreasing trend up to 250 m, while other islands had it up to 350 m. However, correlations were still quite high in all Islands. In this context it was obvious that a single plot with 3 subplots is an effective method to collect data in Indonesian forests. The spatial correlation is decreasing the efficiency of clusters with long distances between plots inside clusters in Indonesian forests. In many other countries spatial correlation is decreasing quite fast and large tracts/clusters are preferred.

2.4 Sample grids in different islands

Sampling grids have been compiled for all Islands separately according to distance parameters. Regular grid and a hexagonal version were compiled for further analysis (see examples in Figure 34). The original density is mainly designed to fulfill the accuracy requirements set for country-wide biomass statistics. When additional and more detailed information needs are specified, sample grids can be intensified.

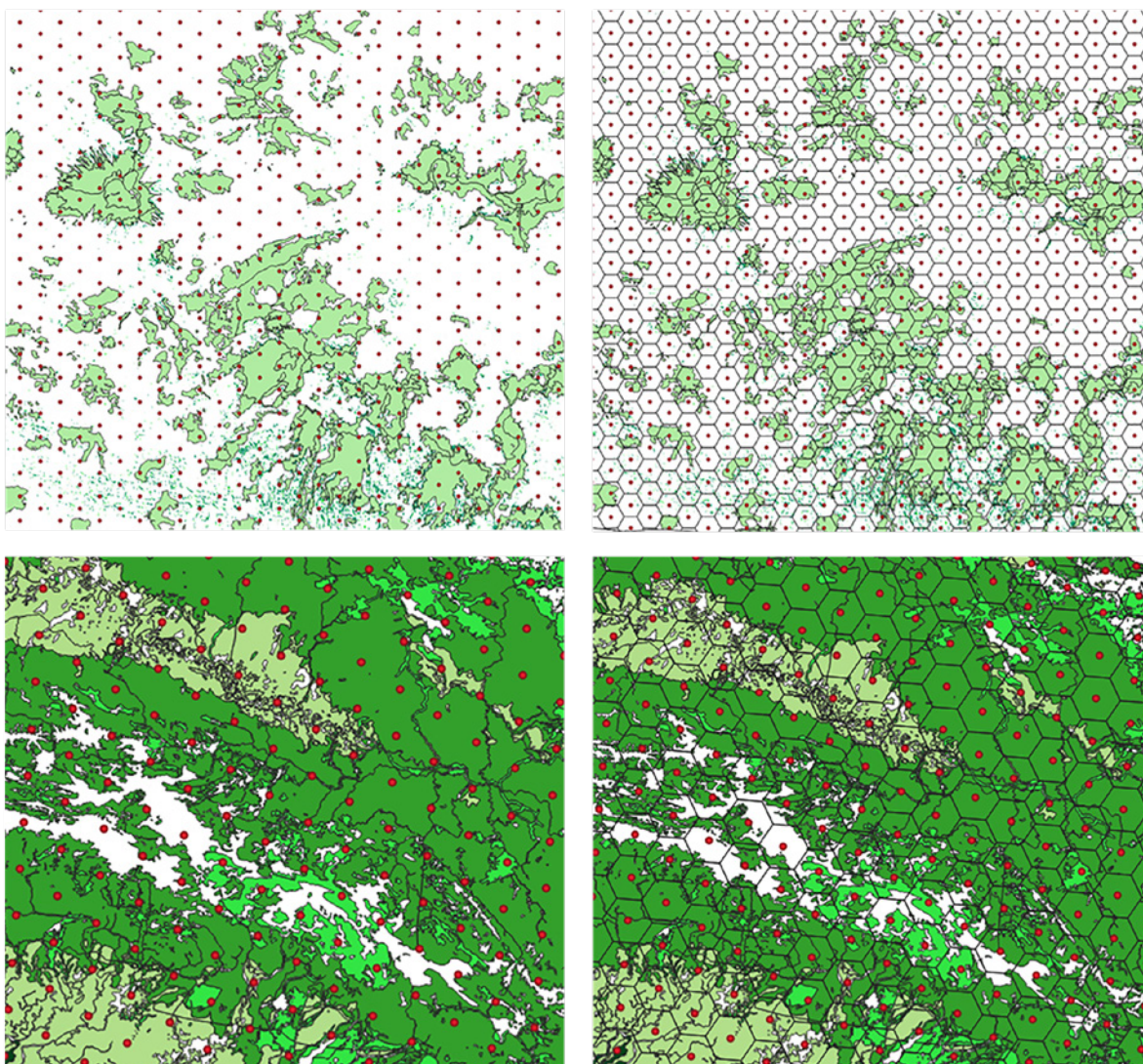


Figure 34. Differences between sparse populations in Jawa (upper part) and dense forest areas in Papua (lower part).

Appendix 3. CORE and OPTIONAL entities and variables

3.1 CORE and OPTIONAL NFI entities and variables

The NFI entities and variables can be split into two categories: core and optional. The core entities' core variables will be collected following field procedures that are consistent and uniform across all areas in Indonesia. The optional variables will be collected either by a subset of sample sites or differently by strata. The following tables describe the variables in data collecting and reporting.

CORE entities:

Trees, Palms, Bamboos, Stumps, Lying deadwood, Canopy parameters (in a subplot)

OPTIONAL entities, e.g.:

Saplings, Seedlings, Shrubs, Climbers, Giant ferns, Other plants, Litter, Soil.

CORE input variables by all entities.

		Tree	Sapling	Palm	Seedling	Shrub	Liana	Bamboo	Stump	Dead wood
Species		x	x	x	x	x	x	x		
Count			(x)		x			x		x
DBH Diameter	cm	x	x	x		x	x	x	x	x
Top height / Length	m	x		x		x	x	x	x	(x)
Bole height	m	x								
Health status		x		x						
Decay status									(x)	x

OPTIONAL input variables are many, for example: Stem form, Stem quality, Tree stump height diameter, soil properties and assessments, soil properties and assessments, etc.

CORE result variables by all entities.

		Tree	Sapling	Palm	Seedling	Shrub	Liana	Bamboo	Stump	Dead wood
Count (i.e.stocking)		x	x	x	x	x	x	x	x	
Mean DBH	cm	x		x			x	x		
Mean height / Length	m	x		x		(x)	x			
Basal area	m ²	x	x	x		x				
Bole volume	m ³	x								
Volume	m ³	x	x	x					x	x
Above-ground biomass	tons	x	x	x		x		x	x	
Below-ground biomass	tons	x	x	x		x		x	x	
Total biomass	tons	x	x	x		x	x	x	x	x
Above-ground carbon	tons	x	x	x		x		x	x	
Below-ground carbon	tons	x	x	x		x		x	x	
Total carbon	tons	x	x	x		x	x	x	x	x
AGB before felling	tons								x	
AGB removal	tons								x	

OPTIONAL result variables are for example: Litter biomass and carbon, Soil organic carbon (by soil depth layers), Biodiversity indexes, Species-area curves.

