A Report to the Food and Agriculture Organization of the United Nations (FAO) in support of Sampling Study for National Forestry Resources Monitoring and Assessment (NAFORMA) in Tanzania

Erkki Tomppo[1], Matti Katila[1], Kai Mäkisara[1], Jouni Peräsaari[1], Rogers Malimbwi[2], Nurudin Chamuya[3], Jared Otieno[3], Søren Dalsgaard[4], Mikko Leppänen[5]

[1] Finnish Forest Research Institute, Vantaa, Finland
PO Box 18 (Jokiniemenkuja 1) FI-01301 Vantaa, Finland
tel: +358 10 211 2170, fax: 358 - 10 - 211 2202
email erkki.tomppo@metla.fi

[2] Sokoine University of Agriculture, the United Rebublic of Tanzania

[3] Forest and Beekeeping Division of Ministry of Natural Resources, the United Rebublic of Tanzania

[4] Food and Agriculture Organization of the United Nations, Dar es Salaam

[5] Food and Agriculture Organization of the United Nations, Rome

10 March 2010

Contents

Acknowledgements

The authors are grateful to the Food and Agriculture Organization of the United Nations (FAO), the Forestry and Beekeeping Division at the Ministry of Natural Resources and Tourism - Tanzania and Sokoine University of Agriculture for this challenging work and for enjoyable and fruitful collaboration in the course of the work. Particularly, we thank Dr. Felician Kilahama, Director of the Forestry and Beekeeping Division at the Ministry of Natural Resources and Tourism of the United Rebublic of Tanzania and Mr. Jim Carle, Leader, Forest Management Team at FAO. We also thank Dr. Inge Jonckheere at FAO for helping in the acquisition of many digital data sets.

Vantaa, Finland, Dar Salaam, Tanzania, and Rome, Italy March 2010

The Authors

1 Introduction

The Food and Agriculture Organization of the United Nations (FAO) has collected global level forest information since 1947. UN-ECE/FAO (United Nations Economic Commission for Europe/FAO) has co-operated with FAO in collecting and publishing the Temperate and Boreal region data (UN-ECE/FAO, 2000). The most recent report, FRA 2005 was published in late 2005 (FAO 2005). Forest resources assessment is based mainly on National Forest Inventories (NFI) conducted at national and sub-national levels.

Countries are planning and conducting the inventories on the basis of their own information needs and traditions. Some countries have long traditions, from the beginning of 1920s, while other countries have conducted just one inventory or are even planning the first sampling based inventory, or at a global level, are lacking even the first inventory. While some inventories are wood production oriented, some other inventories are targeted to produce information about non-wood goods and services, or are multi-purpose inventories. Most inventories collect information on the same base variables; however, some inventories collect some hundreds of parameters measured on the field (Tomppo and Andersson 2008).

During the past 10 years, The Forestry Department of FAO has invested substantial resources to develop a programme of support to national forest monitoring and assessment (NFMA) (Saket et al. 2010). NFMA operates is mainly that of countries in development. Technical, financial and institutional capacity are often needed.

Currently (March 2010), NFA has been completed in nine countries (Bangladesh, Cameroon, Costa Rica, Guatemala, Honduras, Lebanon and Phillipines, Zambia and Nicaragua), is in progress in thirteen countries (Angola, Republic of Congo, Kenya, Kyrgyzstan, Tanzania, The Gambia, Brazil, Comoros, Algeria, Uruguay, Ecuador, Peru and Vietnam) and Formulated for four countries (Cuba, Nigeria, Uzbekistan and Macedonia, FAO 2010).

FAO NFMA has employed a standard approach in sampling design and data collection until year 2009 (FAO 2008). The sampling intensity has been intensified on the basis of local information needs.

FAO NFMA has also launched studies to analyse and further develop the design (FAO 2008). On the basis of these studies, the local information needs and local studies, a somewhat different approach was designed to be used for Tanzania.

This report describes the procedure and results in developing a sampling design for Tanzania forest inventory, NAFORMA. The work was carried out in a tight collaboration with Tanzanian experts, FAO NFMA team and the Finnish Forest Research Institute (Metla) project “National Forest Inventory Designs and Methods” (http://www.metla.fi/hanke/3428/index-en.htm).

1.1 Objectives

On the basis of the agreement between the Food and Agriculture Organization of the United Nations (FAO) and the Finnish Forest Research Institute (Metla), the following activities had to be carried out.

- Further analyze the utility of available remote sensing, digital mapping and inventory data in Tanzania for NAFORMA purposes.

- Study feasibility and provide recommendations on possible alternative sampling and plot designs that could provide Forest Inventory data for Tanzania with improved efficiency and accuracy/precision and that would provide reliable information also at District Level.
- Propose options for efficient sampling designs to meet both the budget and other constraints and the need to make estimates for national and district level. One of the options is the present NFMA approach.
- Analyze and provide recommendations on whether a possible Multi-source Inventory could provide district level Forest Inventory data for Tanzania.

1.2 Description of Activities and Services

On the basis of the agreement, Metla team should do the following tasks.

i Analyze the existing remote sensing material, digital map and inventory data available for Tanzania and how to utilize them in the inventory planning and how to integrate them with field measurements.

ii Propose options for efficient sampling designs to meet both the budget, other possible constraints and the needed to make estimates for national and district level (or for a level of groups of districts in case of small districts) for core SFM/carbon pool and carbon pool change estimates, while at the same time meeting the international reporting requirements. One of the options is the FAO's present conventional NFMA approach. The proposal shall include a proposal for stratification and the sampling designs for each stratum with error estimates for basic parameters, such as area estimates, and volume and/or biomass estimates in case respective functions are available.

iii Study feasibility and efficiency for each sampling options: including accuracy/reliability estimates (for core SFM/Carbon pool and carbon pool change estimates), cost, time, staff and technical skills, transportation and access, and feasibility for repeated measurement applying the same methodology for long term monitoring.

iv Analyze and provide recommendations on whether a possible Multi-source Inventory could provide Forest Inventory data for Tanzania with reliability that would provide information at District Level and further how field plots would be integrated with RS and available digital maps to provide better estimates on core SFM parameters.

The outputs of the project is a sampling Study indicating the analysis of the data, proposed alternative designs and their comparison, stratification plan, and detailed inventory design for Tanzania. Sampling designs for each stratum with error estimates for basic parameters, such as area estimates, and volume and/or biomass estimates should be given, in case respective functions are available.

The deliverables are: 1) assessment of feasibility and efficiency for each sampling options and 2) digital map on location of inventory plots. Furthermore, recommendations for Multi-source inventory should be presented.

2 Principles in a forest inventory and the approach taken

2.1 Basic principles and facts in forest inventory and survey sampling

Let us first recall some basic concepts of in sampling theory and probability sampling, cf. Tomppo and Andersson (2008), Gregoire and Valentine (2008), and Särndal et al. (1992).

1. Target population is a set of the elements for which the inference is to be made. Population can be discrete (finite or infinite) or continuous (always infinite, e.g. a real plane, a forest area).

2. A sample *s* is a subset of the population.

3. Sampling frame is the mechanism which allows to identify the elements in the population.

4. The set of all samples is denoted by *S*.

5. Selection probability of a sample is denoted by *p(s)*. The sampling frame usually determines the selection probability of each sample.

6. Inclusion probability of an element,

$$\pi_i = \sum_{s:U_i \in s} p(s) \tag{1}$$

tells the probability that an individual U_i is included in an arbitrary sample. Note that the inclusion probabilities can vary and that the most efficient sampling procedures often rest upon unequal probabilities (e.g., Mandallaz 2008, see also Gregoire and Valentine 2008).

Probability sampling should satisfy some further conditions.

7. In probability sampling, each element in must have a positive inclusion probability. The inference concerns that set of the elements which have a positive inclusion probability.

8. A sample *s* is selected by a random mechanism under which each possible *s* receives exactly the probability *p(s)*.

Note that each sample *s* does not have to have a positive probability. It is important to take into account the unequal probabilities in the inference. Otherwise, biased estimates may result.

2.1.1 Basic requirements for a forest inventory

The primary traditional purpose of NFIs has been to provide accurate information for forest management and in planning forest industry investments (e.g., Tomppo et al. 2010). In Europe and North America, forest health monitoring emerged as an important issue when increased acidic deposition led to local declines of sensitive forest ecosystems during the 1980s and brought forest health monitoring programmes an essential elements of forest inventories. The role of forests in providing non-wood goods and services such as wildlife habitat, recreational opportunities, and contributions to water quality has received increased attention in recent years, particularly in urbanised societies. Human induced habitat losses, as well as continuing deforestation in the tropics, have resulted in an accelerating rate of species extinction.

The use of fossil fuels, deforestation in the tropics and farming have increased CO_2 and other greenhouse gases (GHG) in the atmosphere and lead to a threat of global warming. Over a decade ago, most countries joined an international treaty - the United Nations Framework Convention on Climate Change (UNFCCC) -to begin considering what could be done to reduce global warming and to cope with the inevitable temperature increases (UNFCCC 2009). More recently, a number of nations approved an addition to the treaty, the Kyoto Protocol (KP), which has more powerful and legally binding measures.

Parties to the UNFCCC process recognized the contribution of GHG emissions from deforestation in developing countries to climate change and the need to take action to reduce such emissions. After a two-year process, the COP adopted a decision on “Reducing Emissions from Deforestation and Degradation in developing countries: approaches to stimulate action” (REDD), Decision 2/CP.13 (UNFCCC 2008). The decision provides a mandate for several elements and actions by parties relating to reducing emissions from deforestation and forest degradation in developing countries.

Examples of the activities are to support and facilitate capacity-building, technical assistance and transfer of technology relating to methodological and technical and institutional needs of developing countries, and to explore demonstration activities to address drivers of deforestation and enhance forest carbon stocks due to sustainable management of forests.

The changing roles of forests and the requirements for national reporting and common international reporting have substantially altered demands for forest information. Because NFIs are the primary source of forest information for all these purposes, the scope of NFIs has broadened accordingly and has resulted in the introduction of a wide variety of new variables requiring assessment.

Despite these new ways of using NFI data, one of the most important roles of NFIs is as the key information source for national forest policies and forestry programs by providing a picture of the status of the forests at sub-national and national level.

The forest inventory requirements can be summarised as follows: To provide information to asses

> Current status and changes in forests and to take into the existing and coming multidisciplinary needs, such as the information for
>
> - forestry and strategic planning of forestry
> - different timber uses
> - assessing and maintaining forest biodiversity
> - forest assessing forest health status
> - assessing the role of the forests in carbon balance
> - assessing the possibilities to reduce emissions from deforestation and degradation (REDD).

The inventory design should take into account these needs, and further fulfil the efficiency requirements:

- each observation should bring new information
  - more measurements on areas where variation or changes are high
    - stratification is one tool to meet efficiency requirements

Examples of the other requirements and facts are:

- the entire land area should be covered
- the reporting units and the output parameters should be defined
- international definitions, taking into account local conditions should be used
- accessing each measurement unit often requires lot of time
- financial constraints.

One of the key questions is: how to determine the spatial layout (design) of the sampling units?

2.1.2 The approach taken in the study

The approach taken was sampling simulation for error estimation and GIS analysis based cost assessments. The ideas presented in Tomppo and Katila (2008) were used. For this, a model of land use and/or land cover with the road network, elevation variation of the terrain and volume of growing stock are needed.

The input data sets and the methods in preparing them are first described, after that, the selected statistical framework for the design, the simulation procedure, and finally the results. Note that the main principles of the coming result calculation system have also been developed.

3 Input data sets for the sampling study

For the sampling simulation, a spatial model, i.e., digital geo-referenced data sets, of the land use and forests is needed. Using this model, optional sampling designs can be picked up and the efficiencies of the designs compared through calculating the sampling errors of the basic parameters as well as the measurement costs of the field plots for each design. The spatial model was prepared by means of the available land use and vegetation cover data as well as the predictions of the relevant forest variables which variables were predicted in the study. The first task was to identify the necessary and available and data sets.

The following data sets were employed in the study.

1. Landsat ETM+ satellite image mosaic over Tanzania
2. The MODIS Surface-Reflectance Product (MOD 09) over Tanzania

3. Hunting map over Tanzania, vegetation types, roads, etc.
4. Digital elevation model
5. Aggregated forest inventory data from eleven Districts, Tanzania
6. District boundaries from Tanzania
7. Africover map database (URL `http://www.africover.org/`), rivers
8. Field plot data from East Finland
9. Landsat ETM+ satellite images from East Finland
10. The MODIS Surface-Reflectance Product (MOD 09) over East Finland

All the data sets, except Finnish data, were converted into the map projection UTM (zone 36 South) with the WGS84 datum.

4 The Landsat Image Data

4.1 Image Selection and Preprocessing

Three sets of images were originally considered as basis for image data. They were GLS 1990, GLS 2000, and GLS 2005 (Global Land Survey). All sets have been created by U.S. Geological Survey (USGS) and NASA. The 1990 data set consisted of Landsat TM images and the 2000 data set consisted of ETM+ (Landsat 7) images. The 2005 set is a combination of Landsat TM and ETM+ images. The GLS 2000 data set was selected for this project. The image quality over Tanzania seemed better than in the 2005 data set. The year 2000 was also closer to the available ground data than 2005.

The Landsat data set consisted of 59 images. Visual inspection revealed that three of the images were too hazy to be used (the automatic cloud detection algorithm used by USGS did not find the haze). Substitutes were selected for these images (WRS-2 indices path-row: 172-62, 172-63, and 172-64). The images were from years 1999 - 2002. There were images from all seasons of the year at different parts of the country. A list of the images is in Table 1.

The regions of the images covered by clouds or cloud shadows can not be used in predicting forest variables. Cloud and cloud shadow masks were manually made for the images using an image display.

The Landsat images obtained from USGS were already orthorectified. The map projection used in the images was UTM with the WGS84 datum. The images over Tanzania were in four UTM zones (34, 35, 36, and 37). The zone 36 (South) was selected to use in processing over Tanzania because it covered the largest area. Using one zone simplified processing because a single image mosaic covering the whole country could be used. The pixels size chosen for processing was 30 meters (in the UTM 36S system).

The mosaicing was done by overlaying images one over another. The overlaying order was selected to prefer better quality images. There is not much overlap in the Landsat images of adjacent WRS paths. Because of this, the mosaicing order did not enhance the result much. The first mosaic made from the original image data transformed to reflectance at the top of the atmosphere is seen in Fig. 1.

In addition to the Tanzania data set, two adjacent Landsat 7 ETM+ images from North-Karelia in Finland was used for development of the volume model (WRS-2 path-row: 186-16, 186-17, 2000-06-10).

The image processing was performed mostly using in-house software made at Metla. The open source software GDAL (URL `http://www.gdal.org/`) was used in changing image formats and transform the data between UTM zones. All image processing was done using 64-bit Linux systems.

Table 1: The Landsat images over Tanzania used in this project. The image later in the patch order overwrites earlier data in the mosaic.

WRS-2			Patch		WRS-2			Patch	
Path	Row	Date	order	Comment	Path	Row	Date	order	Comment
165	65	2001-09-14	11		169	63	2000-02-12	6	
165	66	2000-05-22	38		169	64	2000-02-12	7	
165	67	2000-05-22	39		169	65	2000-02-12	8	
165	68	2002-05-12	12		169	65	2000-07-21	9	
166	63	2000-01-22	40		169	66	2000-07-21	10	
166	64	2000-01-22	41		169	67	2001-09-10	4	
166	65	2000-06-30	42		170	61	2001-05-12	51	
166	66	2000-06-30	43		170	62	2001-05-12	52	
166	66	2001-05-16	13		170	63	2000-01-18	27	
166	67	2000-05-29	14		170	63	2001-07-15	26	
166	68	2000-06-30	15		170	64	1999-10-14	28	
167	62	2001-03-04	44		170	64	2000-01-18	29	
167	63	2001-03-04	45		170	65	2000-12-19	31	
167	64	2001-01-15	16		170	65	2001-10-03	30	
167	65	2000-07-07	17		170	66	2000-12-19	32	
167	66	2002-05-10	18		170	67	2001-08-16	33	
167	67	1999-10-25	19		171	61	2000-07-19	1	
167	67	2000-07-07	20		171	62	2000-05-16	53	
167	68	2001-12-01	21		171	63	2000-05-16	54	
168	61	2000-02-21	46		171	64	2002-05-22	55	
168	62	2000-02-21	47		171	65	2002-05-22	56	
168	63	2000-02-21	48		171	66	2002-05-22	57	
168	64	2001-03-11	22		172	61	1999-07-08	2	
168	65	2000-02-21	23		172	62	2002-05-13	34	not GLS 2000
168	66	2002-06-18	49		172	63	2002-05-13	35	not GLS 2000
168	67	2000-02-21	24		172	64	2002-05-13	36	not GLS 2000
168	67	2002-06-18	50		172	65	2002-05-13	37	
168	68	2000-08-31	25		173	63	1999-09-17	58	
169	61	2000-01-27	3		173	64	1999-09-17	59	
169	62	2000-02-12	5						

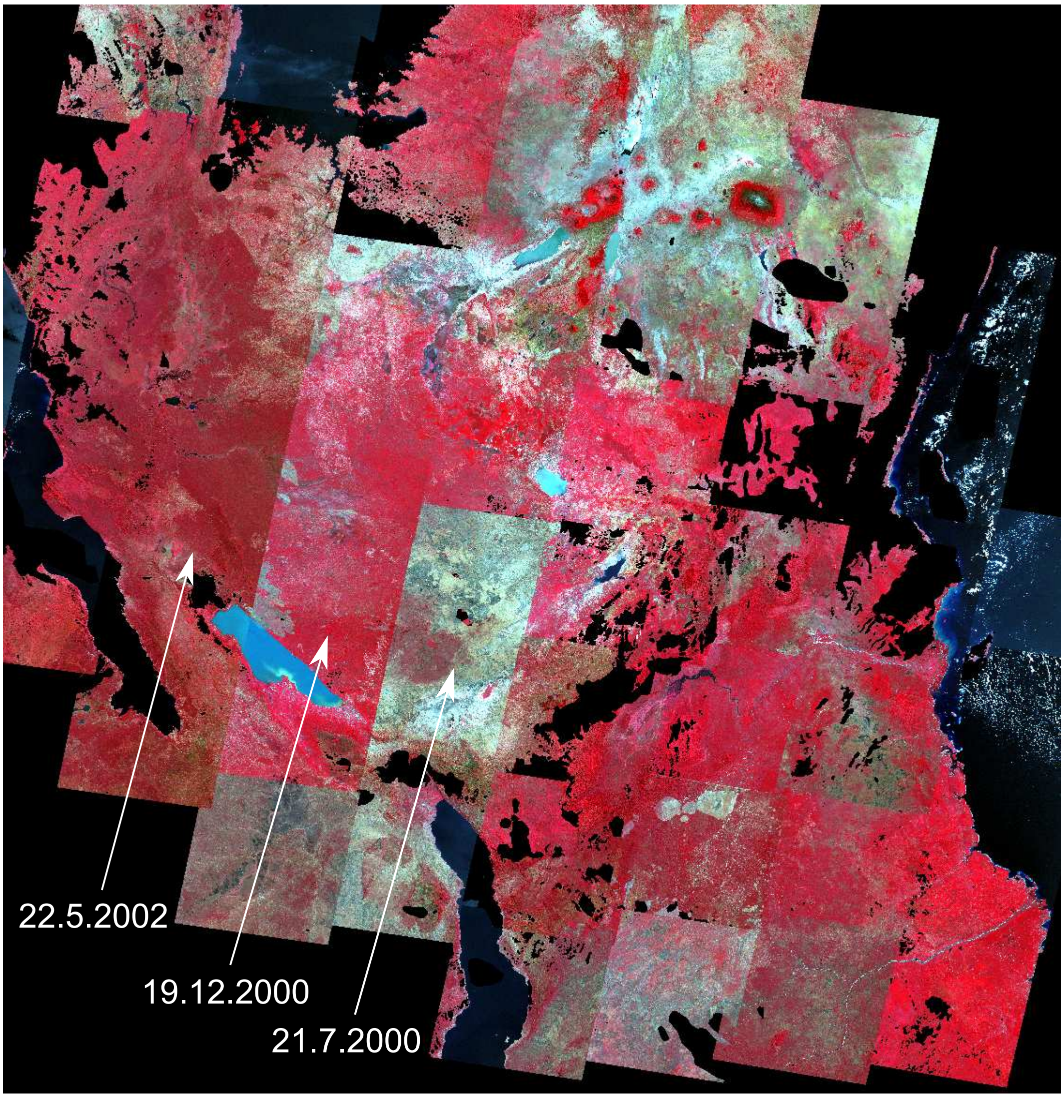

Figure 1: The Landsat image mosaic over Tanzania made from the original image data transformed to top of atmosphere reflectance. The spectral channels shown are 4, 3, and 2 (Red, Green, Blue).

4.2 Atmospheric Correction

The first mosaic was made from the original image data transformed to reflectance over the top of the atmosphere. This processing removes the effects of the different imaging dates and times but does not do anything to compensate for other factors. The effects caused by the different imaging conditions (atmosphere, angular dependence of the target reflectance, etc.) and seasonal differences in vegetation were clearly visible in the mosaic (see Fig. 1).

This mosaic would make prediction of forest variables very difficult. To simplify prediction, an atmospheric correction of the image data was performed. The method used an atmospherically corrected image with large pixel size covering the whole area of the mosaic. The image data of the individual Landsat images were then adjusted to match this coarse image.

The image data produced by the MODIS instrument (URL `http://modis.gsfc.nasa.gov/`) of NASA was used to provide the reference image. MODIS is currently flying on two platforms: the TERRA and the AQUA satellites of the NASA EOS (Earth Observing System). Several image products are made operationally from the image data. One of these is the surface reflectance (MOD09, URL `http://modis.gsfc.nasa.gov/data/dataprod/dataproducts.php?MOD_NUMBER=09`). This product is computed from the original data using a very advanced atmospheric correction method. This product is distributed in several formats. The format chosen for this project is a mosaic from the best data over 8 day periods (MYD09A1). The pixel size in this format is 500 meters.

Using another image as reference for atmospheric correction requires that both the image corrected and the reference image include compatible spectral channels. This is the case for Landsat channels 1, 2, 3, 4, 5, and 7 (MODIS channels 3, 4, 1, 2, 6, and 7). The MODIS data MYD09A1 is provided in the sinusoidal projection in 10 by 10 degree tiles. Three tiles covering the mosaic area (h20v9, h21v9, and h21v10) were acquired and transformed into the UTM projection using the Modis Reprojection Tool (URL `https://lpdaac.usgs.gov/lpdaac/tools/modis_reprojection_tool`). In addition to the image data, the MODIS product included also a usable cloud mask

Good MODIS data for this purpose was surprisingly difficult to find. The time interval of the data should be somewhere between January and April. This is due to fact that the normalization is done to the growing season conditions. The data should be as cloud-free as possible. Taking these things into account, the best data found was from the MODIS instrument on AQUA satellite from February 26, 2003.

The North-Karelia image was also atmospherically corrected. In this case, MODIS AQUA image from July 4, 2002 was used (tile h19v02).

The atmospheric correction was computed separately for each Landsat image before mosaicking. The correcting function was a simple linear mapping

$$y_i = a_i x_i + b_i, \tag{2}$$

where x_i is the uncorrected data for channel i, y_i is the corrected data, and a_i and b_i are parameters computed for each channel. The basic principle is to match the mean and the variance of the data in both images taking into account the different pixels sizes. To do this, the Landsat data is transformed to 500 meter pixel size by averaging over blocks of 16 by 17 pixels. The parameters a_i and b_i are computed by

$$a_i = s_{mj}/s_{li} \tag{3}$$
$$b_i = m_{mj} - m_{li}/a_i \tag{4}$$

where m_{li} and s_{li} are the mean and standard deviation of Landsat channel i and m_{mj} and s_{mj} are the mean and standard deviation of the compatible MODIS data channel. The means and standard deviations are computed over the pixels that are cloud-free in both images.

The atmospherically corrected mosaic is seen in Fig. 2. The result shows that the atmospheric correction is fairly efficient but not perfect. Using image-wide correction parameters does leave within-image variations in the data. The method does not include any reasonable correction for the seasonal effects.

There may or may not be some novelty in the method. Exactly this method has not been reported in literature, but some indications have been seen that other people have developed methods of similar type.

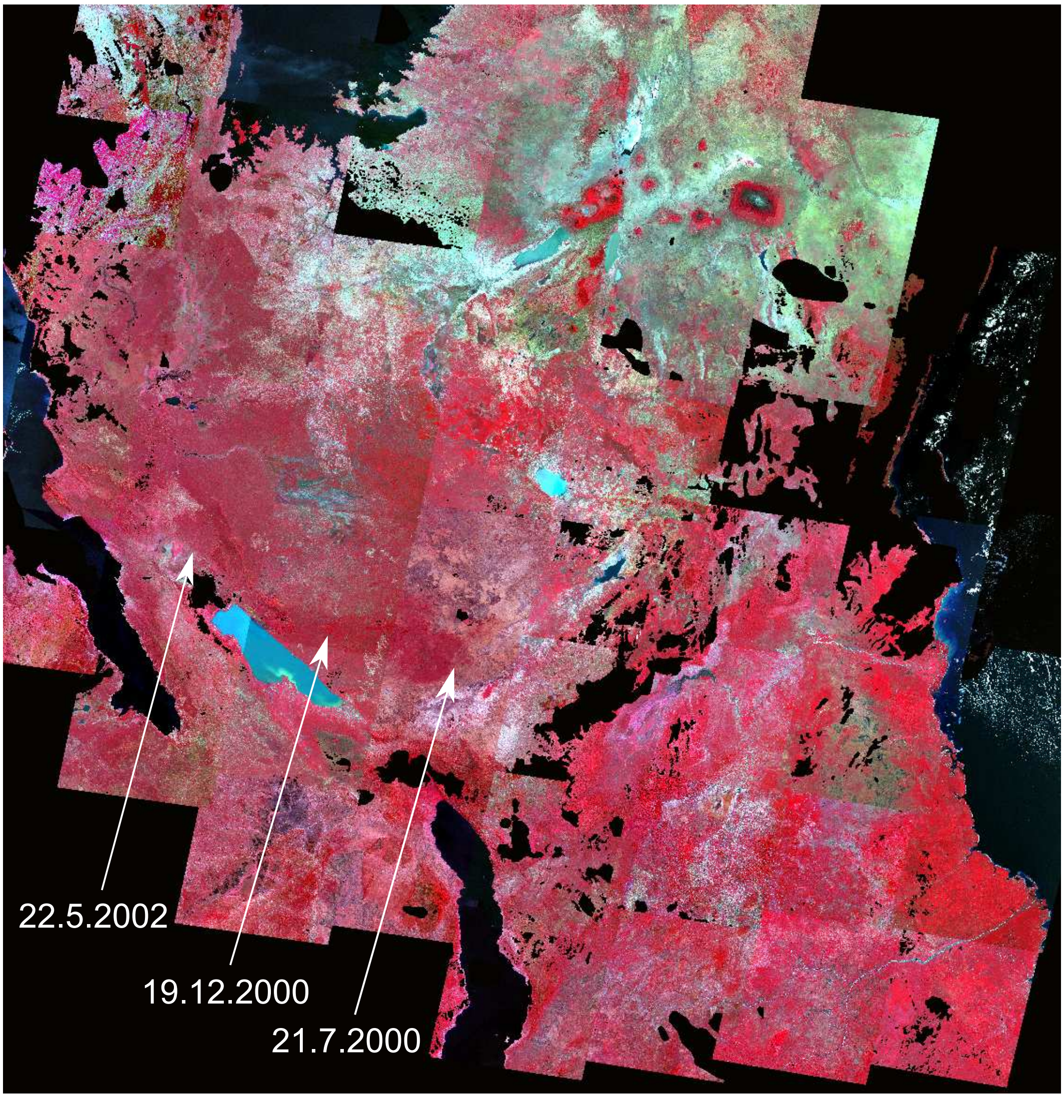

Figure 2: The atmospherically corrected Landsat image mosaic over Tanzania. The spectral channels shown are 4, 3, and 2 (Red, Green, Blue).

5 Map data

Hunting map sheets were down-loaded from the FAO website. The selected themes were the vegetation maps and roads. The map sheets were combined and projected to UTMs 36 system with the WGS84 datum. Originally the sheets were in UTM 36 and 37 projection with ARC 1960 ('ARS_B' in Arc/Info software) datum. All the map themes were rasterized to 30 m pixel size. The vegetation types were coded according to the 'vegnbr' number in Table 2. A re-classification of vegetation types to 7 strata ('Prestrat' in Table 2) was done and a corresponding raster layer produced (Fig. 3). The aim of the re-classification was to form strata large enough for sampling simulation purposes and to define land classes (forest, wooded land) for which the within class growing stock density was on the average similar and for which the sampling simulation results had to be be calculated. The average growing stock was estimated for each vegetation type, based on expert opinion, and the vegetation types were regrouped to the seven strata using this information. The Hunting road classes 1–5 (roads and 'footpaths') were rasterized. Laid over the Landsat images, it was found out that there are locational errors in the road data ranging from 100 to 300 m. From Africover database, the rivers were taken and rasterized to a grid. The Political boundaries and the District boundaries were also taken from Africover database. The Africover (URL `http://www.africover.org/`) map data were originally in geographic projection with the WGS84 datum. The average walking speed raster layer was produced from the Hunting vegetation map using the estimates given for different vegetation types in the Table 2. The rasterized roads and rivers were overlaid on the walking speed raster with values 10 min/km and 200 min/km, respectively. In this way, the barriers in the walking paths caused by major rivers were to be taken into account.

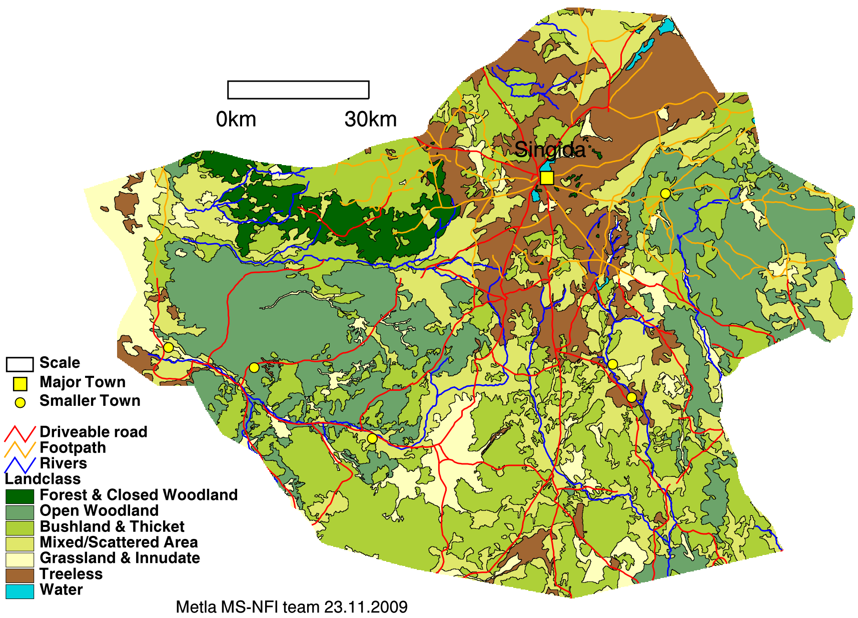

Figure 3: Re-classified vegetation types based on Hunting map, Singida District.

The vegetation types belonging to the "Pre-strat" codes 1-3 are called in the following "Forest land" although the stratum is not necessarily compatible with the international definition of Forest. Similarly, the stratum consisting of the vegetation types with "Pre-strat" codes 1-6, excluding "Vecodes" 42, 43, 52, 64, 70, 71, 72, 73 and 80, is called in the following "Wooded land".

A digital elevation model (DEM) with pixel size of 90 m origins from SRTM data (URL `http://eros.usgs.gov/\#/Find\_Data/Products\_and\_Data\_Available/SRTM`). The DEM data were in geographic projection with the WGS84 datum and projected to UTM 36 South system with the WGS84 datum. There were problems in the use of DEM in calculating the slope of terrain with the algorithm employed due to the much smaller output pixel size, 30 m, than in the original data. Extra stripes appeared in the slope image. A linear filtering (lowpass) was therefore carried out for the DEM. A window size of 15×15 pixels was used. The Gaussian weight function (Eq. 5) was used.

Table 2: Hunting map vegetation classes, pre-stratified land classes 1-7, walking speed assumptions and average plot measurement times for vegetation classes.

Vegcode	Vegnbr	Pre-strat	Description of vegtype	walk speed, min/km.	plot meas. min
Fn	10	1	Natural Forest	60	40
Fm	11	1	Mangrove Forest	40	40
Fp	12	1	Plantation Forest	20	40
Wu	23	2	Woodland (unspecified density)	30	30
Wc	20	1	Closed Woodland	30	30
Wo	21	2	Open Woodland	30	30
WSc	22	3	Woodland with Scattered Cropland	30	30
Bu	30	3	Bushland (Unspecified Density)	30	25
Bd	31	3	Dense Bushland	30	25
Bo	32	4	Open Bushland	15	25
BSc	33	4	Bushland with Scattered cropland	30	25
B(et)	34	5	Bushland with Emergent Trees	30	25
Bt	35	3	Thicket	40	40
Bt(et)	36	3	Thicket with Emergent Trees	40	40
Gw	40	4	Wooded Grassland	15	25
Gb	41	5	Bushed Grassland	30	25
Go	42	6	Open Grassland	15	15
0GSc	43	6	Grassland with Scattered cropland	25	15
Gws	50	5	Wooded Grassland (Seasonally inundated)	25	25
Gbs	51	5	Bushland Grassland (Seasonally inundated)	25	25
Gos	52	6	Open Grassland (Seasonally inundated)	25	15
Cm	60	3	Mixed Cropping	25	20
Ctc	61	2	Cultivation with Tree crops	25	20
Ctc(st)	62	2	Cultivation with Tree crops (with shade trees)	25	20
Cbc	63	5	Cultivation with Bushy Crops	25	20
Chc	64	6	Cultivation with Herbaceous crops	25	15
BSL	70	6	Bare Soil	30	10
SC	71	6	Salt and Crusts	40	10
RO	72	6	Rock Outcrops	40	10
ICE	73	6	Ice cap - snow	200	10
Ocean	91	7	Ocean	200	0
IW	90	7	Inland Water	200	0
S/M	54	5	Swamp/Marsh (Permanent)	100	15
Ua	80	6	Urban Areas including air fields	10	10

$$f(x,y) = ae^{-\frac{x^2+y^2}{2c^2}} \tag{5}$$

where x and y are pixels (planar) coordinates. The value $\sqrt{2}$ was used for parameter c (Fig. 4 and 5). The slope in percentage values (45° is equal to 100 %) were calculated from filtered DEM using the cross-pixels in a 3×3 window.

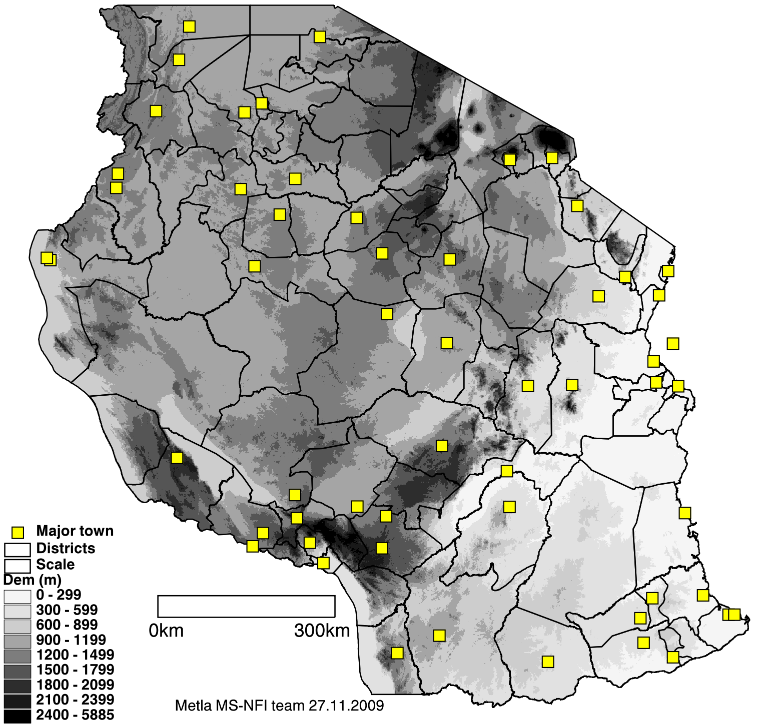

Figure 4: Input data, DEM.

The calculation of the Euclidean distance and Pathdistance from the road network is explained in section 7.1.

6 Volume estimation and calibration

The map form predictions for volume of growing stock were employed in analysing the cost-efficiency of different sampling designs. Plot level field data from Finland were used due to the lack of plot level data in Tanzania. A robust model was estimated and employed. It was assumed that the ratios of the spectral values of different Landsat ETM+ spectral bands are are not as sensitive to the change of the vegetation zones as are the absolute reflectance values. Other variables, such as brightness, greenness, wetness, were also tested.

The parameters of the were estimated using Finnish data, top of atmosphere reflectance, with atmosphere calibration, and non-linear estimation

The number of the explanatory variables was intentionally kept low the preserve the robustness. The selected model was

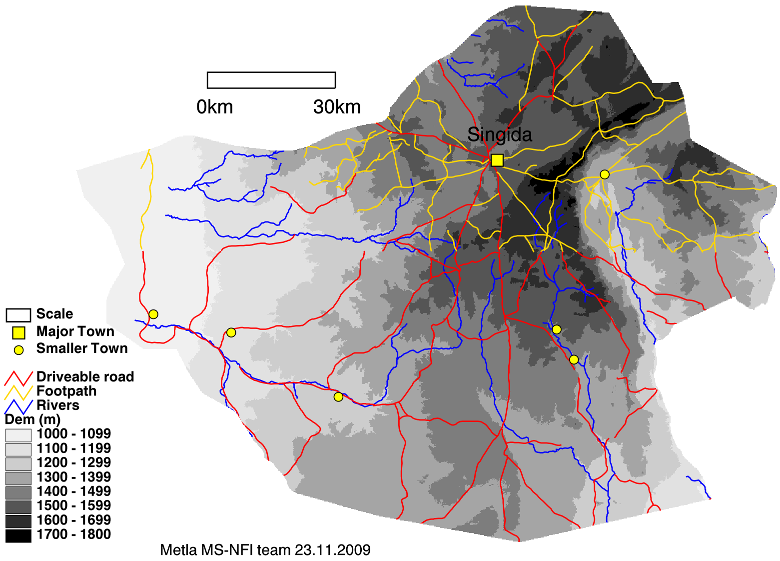

Figure 5: Input data, digital elevation model, DEM, Singida District.

$$vol = c * exp(a + b_1 * u_3/u_2 + b_2 * u_7/u_5) + \varepsilon \tag{6}$$

where u_i is the reflectance of the Landsat ETM+ spectral band i as given in section 4.2 and c, a and b_i the parameters of the model.

The parameters of the model were estimated using non-linear regression and SAS NLIN procedure. The model explained 75 % of the volume variation of the volume.

The model were calibrated using the aggregated data from 11 districts (see Fig. 6). The final model after the calibration is

c=1.2146, a=15.943, b_1=-29.3802 and b_2=3.2762

The volumes for the land areas covered by the clouds or cloud shadows were predicted as follows. The averages and standard deviations of volume predictions were calculated by the Hunting map categories (Table 2). The prediction for each pixel under the cloud or cloud shadow was randomly drawn from the normal distribution with the mean and standard deviation equal to the empirical mean and standard deviation of the volume predictions for the corresponding Hunting map category.

The volume predictions after the calibration is shown for entire Tanzania in Fig. 7 and for Singida District in Fig. 8.

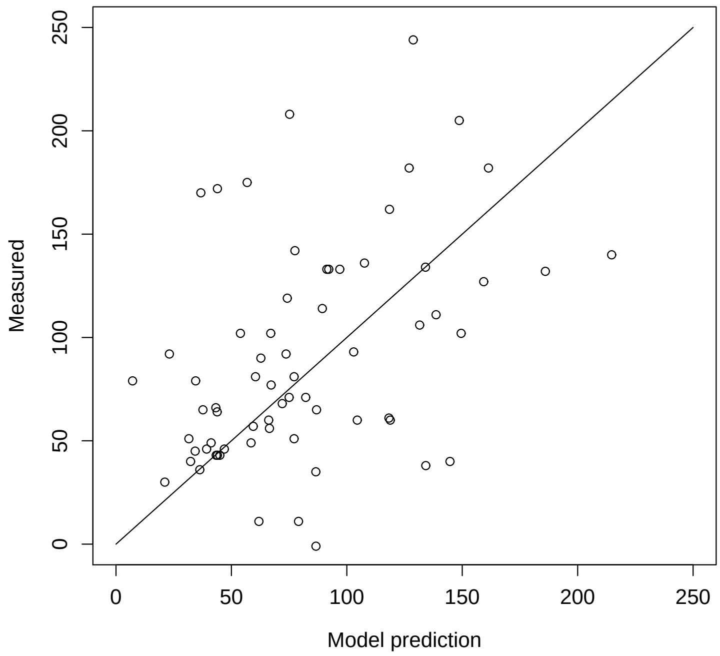

Figure 6: The field data based estimates (Measured) of the volume of growing stock (m^3/ha) against the estimates of the volume of growing stock based on the model using Finnish field plot data and Landsat ETM+ 7 images (Model prediction).

7 Time consumption in the field work of different sample designs

In this project, the time used by a field crew to measure a field plot or a cluster of field plots was divided to several phases. The time (minutes) needed for each phase in the field work was based on information given by the Tanzanian participants of the project: walking speed on different vegetation types, assumed time needed to measure a field plot on different vegetation types. Also, the available GIS data and software were used to estimate distances and cost of reaching clusters from the road.

Several assumptions were made to get the time costs of simulated clusters: The walking speed with GPS device in the field and the daily lunch break in the field were assumed to be independent of the design selected. The lodgment is to be ’on-the-road’ camping. A daily pause of 60 min was included, this includes so called ’other actions’ on field apart from measuring the plots.

For systematic cluster samples of L-shape, walking speed assumptions and average plot measurement times presented in Table 2 were used. These values were broken down to ’Hunting map’ vegetation classes. The walking speed from the road to the plot is equal to the speed along the tract (between plots in a cluster) if the terrain is similar because the GPS equipments are used to locate the plots.

Summary of components considered in time calculations:

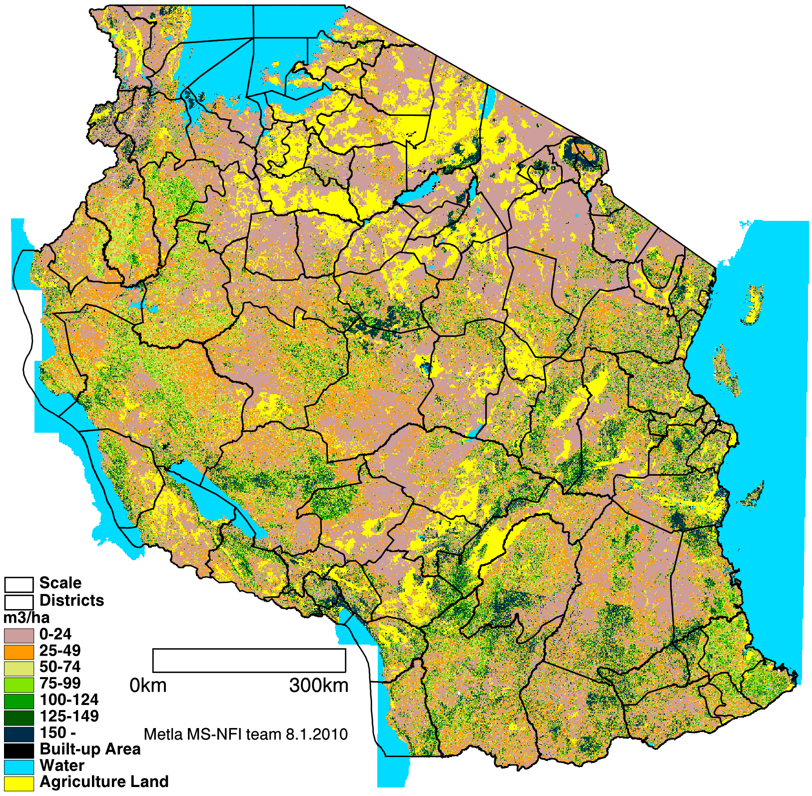

Figure 7: Predicted volume of growing stock.

- Driving to a cluster from the lodgment (50 min)
- Walking in the field (with GPS) to a cluster and along the cluster, walking speed depending on the Hunting map vegetation class
- Measurement of a plot, estimated time per plot according to Hunting map vegetation class
- Daily pause: lunch break and ’other actions’ on field (60 min)

7.1 The driving times, distance from road to field plots, walking on the field

Only one way driving time to the first plot or cluster was included into the measurement time (working hours, 480 min). An average driving time of 50 min from the field camp to the cluster of the field crew were used.

The distance from the road to the field plot was estimated using the ’Hunting map’ road network (Hunting road classes 1-5, including ’Footpaths’). It seemed that, in some areas, a part of the roads were missing. Therefore, the inclusion of the ’Footpath’ road class to the road network was considered reasonable. The Euclidean distance in geographical horizontal space from the nearest road point to the field plot or to the field plots in a cluster was calculated (Fig. 9 and 10). However, the estimate of time used to reach the cluster was calculated using Arc/Info ’Pathdistance’ function, see URL `\smallhttp://webhelp.esri.com/`

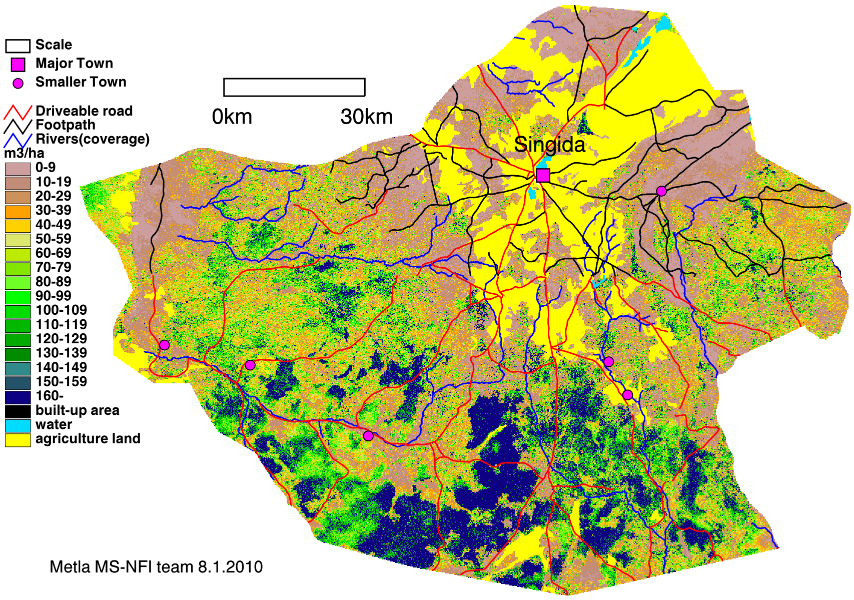

Figure 8: Predicted volume of growing stock, Singida District.

arcgisdesktop/9.2/index.cfm?TopicName=Path_Distance\%3A_adding_more_cost_complexity

Pathdistance determines the minimum accumulative travel cost from a source (in this case the Hunting roads) to each cell location on a raster (a field plot). In Path Distance, we used accumulative cost over a cost surface (walking speeds) defined for different Hunting map vegetation classes (Table 2), the elevation model to compensate for the actual surface distance that must be travelled and the vertical factors (travel time due to the different walking speed associated with downhill and uphill movements). All these data affect the total cost of moving from one location to another.

The formula from Aitken (1977) and Langmuir (1984) (based on Naismith's rule for walking times) was used to estimate the vertical cost factors of different slopes (Eq. 7).

$$T = a\Delta S + b\Delta H_{uphill} + c\Delta H_{moderate\ downhill} + d\Delta H_{steep\ downhill}, \tag{7}$$

where T is time of movement in seconds, ΔS is the distance covered in meters, ΔH is the altitude difference in meters. Parameter a is 1/walking_speed, the average used for Tanzania, 1.8 s/m (from 30 min/km, see Table 2). Parameter values b, c, d (6.0, 1.9998, -1.9998) were used, (moderate downhill $5^\circ < \alpha < 12^\circ$, steep downhill $\alpha > 12^\circ$). α is the angle of the slope, in the Pathdistance function the angle between two cells calculated from the elevation model using the Pythagorean theorem. For details, See, URL http://www.grass.itc.it/grass62/manuals/html62_user/r.walk.html).

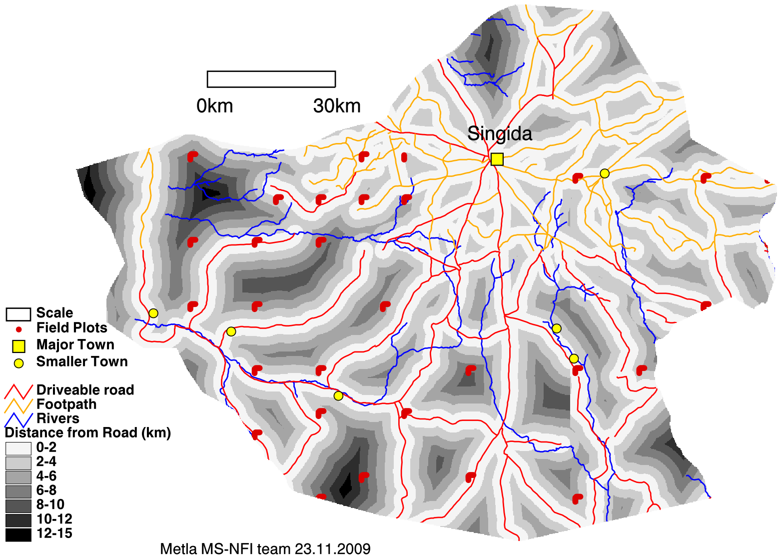

Figure 9: Input data, distance to the nearest road, Singida District.

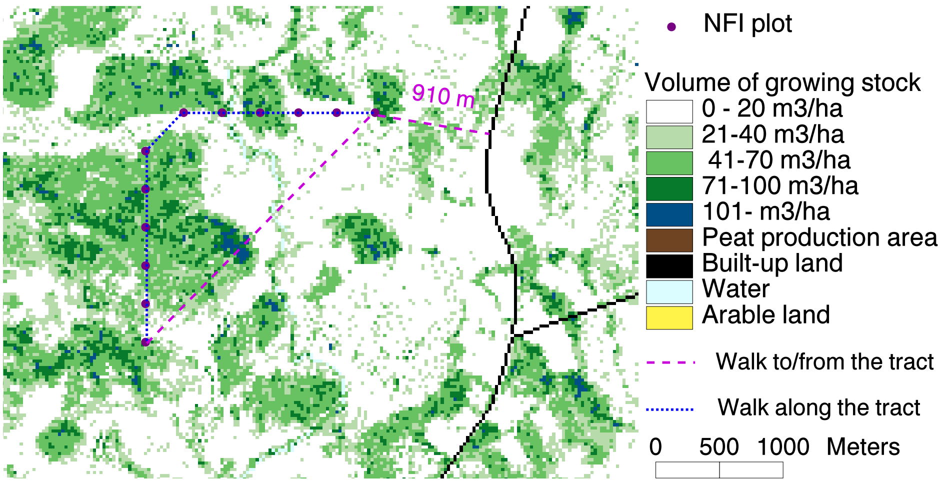

Figure 10: The minimum Euclidean distance (910 m) from the road to the closest field plot of a NFI cluster. The mean volume of growing stock of Finnish multi-source NFI9.

The time estimates T_α for different angles α were related to the one on flat terrain, T_0: The vertical factors VF were calculated as an average of both up- and downwards direction (The same path was assumed to be used for return) $VF_\alpha = (T_\alpha + T_{-\alpha})/2T_0$. The final Vertical factors used with Pathdistance function are presented in Table 3. Beyond the range of these slope values the VF was set to infinite (inaccessible) in Pathdistance analysis.

Table 3: Vertical factors used with Pathdistance function.

Slope °	Vertical factor	Slope °	Vertical factor
-50	3.65	3	1.09
-40	2.86	4.99	1.15
-30	2.28	5	1.10
-20	1.81	8	1.16
-15	1.60	11.99	1.24
-12	1.47	12	1.47
-11.99	1.24	15	1.60
-8	1.16	20	1.81
-5	1.10	30	2.28
-4.99	1.15	40	2.86
-3	1.09	50	3.65
0	1.00		

Only field plots on land were considered. For the L-shaped NFI clusters, the walking distance between the field plots was the distance along the tract line between the two furthermost field plots on land plus the Euclidean distance between them, i.e. the length of the sides of a triangle. Along the tract line (the legs of the triangle) the walking speed was estimated of walking speed on plot points (depending on the Hunting vegetation type) multiplied with VF from Table 3. An average of these speed estimates was taken for the cluster. Along the 'hypotenuse' the walking speed was estimated from ratio of differences of maximum and minimum of pathdistance (min) and Euclidean distance (m) among the cluster plots i, $(max(pathdist_i) - min(pathdist_i))/(max(eucdist_i) - min(eucdist_plot_i))$.

A coefficient of 1.1 was used to multiply all the walking distances or walking time estimates to approximate the need to avoid water areas and other obstacles. The time needed to reach possible plots on small islands was not considered separately in the calculations.

7.2 The time cost for a cluster

The total daily working time is 480 minutes. For the cluster sampling designs the total amount of days needed to measure all the plots in a sample are calculated in such a way that the clusters are divided into two groups, 1) those which take either less than 350 minutes or more than 480 minutes, and 2) those which take more than 350 minutes but not more than 480 minutes. For the former ones, the total working time in days is obtained dividing the total minutes by 480. The latter clusters are considered to need one day each, i.e. a field crew will not continue to another cluster on that day. If the time needed exceeds 480 minutes, a crew will continue measuring the cluster, staying on field, on the next day.

Using the above assumptions, GIS data and analysis, the costs in time (minutes) for each cluster, were calculated for a systematic grid of plots. The simulation programme called a subroutine which returned the time cost for the cluster in question. The cost estimates can then be used in the allocation of the sample plots in each stratum (see Double sampling for stratification) as well as to estimate the total cost of a particular sample.

8 Design

8.1 Elements of a design

A sampling design for a forest inventory is affected by several factors. Examples are

1 Sample plot size

2 Sample plot shape

3 Spatial layout of the plots

- Detached plots / plot clusters
- Distances between the plots
- Distances between possible clusters

4 Stratified design / non-stratified design

- definition and number of the strata
- the questions in point 3 for each stratum in case of stratified design

The solution is far from trivial and depends also on the parameter in question, e.g., area estimate, volume estimate, estimate of rare events. Practical things must be taken into considerations, the measurement unit should be a work-load of one day for a field crew. In theory, the optimal design could be sought minimising the standard errors of the parameters with the fixed costs, or minimising the costs with a given standard errors. In case of only one parameter of interest, e.g., the mean or total volume of growing stock, a stratified sampling design with an optimal allocation could be one solution (Cochran 1977). In practice, a unique solution does not exist because different parameters may require different designs.

The field plot shape and size was agreed in a collaboration with the Tanzanian experts. The plot size was based on the studies carried out earlier by Professor Malimbwi, Dr. Zahabu and their teams at Sokoine University. A concentric plot with a maximum radius of 15 m can be argued for Tanzania conditions (Fig. 11).

8.2 Variogram and semivariance, tools to assess plot distances

Detached clusters of the plots instead of detached individual plots can be argued using the cost-efficiency criteria. These types of designs are currently employed most of the European countries (Tomppo et al. 2010).

Relevant questions are the distances between the plots in a cluster and distances between the clusters. The distance between the plots should be high enough in order that each field plot will bring information. On the other hand, however, the distance always increases the walking time between plots and also the measurements costs. A trade-off between new information and measurement costs should must be adopted. The between plots distance could be studied in terms of geostatistical concept called a variogram. The variogram

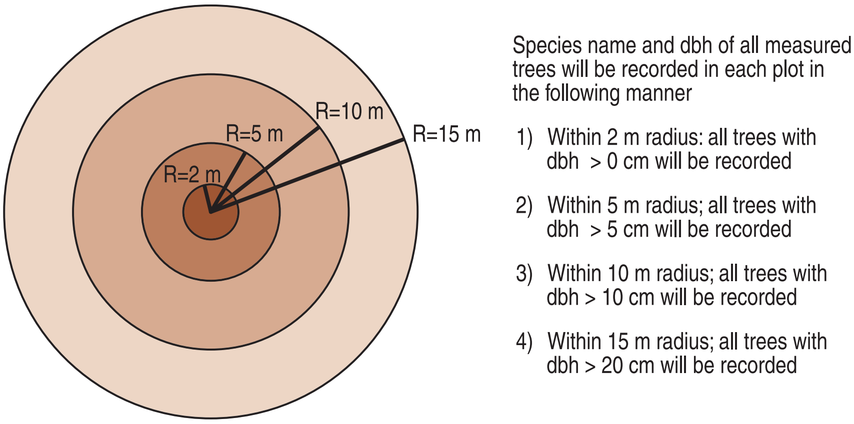

Figure 11: A concentric field plot in cluster designs, max radius 15 m.

of a stochastic process in the two-dimensional plane, Z, e.g., mean volume of the growing stock is defined as a variogram. A robust estimate, semivariance, is defined as (Cressie 1993)

$$2\gamma(x,y) = E\left(|Z(x) - Z(y)|^2\right) \tag{8}$$

where x and y are arbitrary points in the plane.

Many estimates are available for a variogram. A robust estimate, semivariance, is defined as (Cressie 1993)

$$\hat{\gamma}(r) = \frac{(1/N(r)\sum_{S(r)}|x-y|^{1/2})^4}{0.914 + 0.988/N(r)} \tag{9}$$

The semivariances calculated from the volume predictions of the re-classified vegetation types of Hunting map, for “Pre-strat” categories 1-4 (Table 2), are shown in Fig. 12 and for distances 0-1000 meters in Fig. 13. The semivariances level off at the distance of about 250 meters and thus support that field plot distance between the field plots on a cluster.

8.3 The field plot cluster shape and size

Possible shapes of the field plot clusters are for example a rectangle, an L-shape cluster, a line, a hexagon etc. An advantage of a closing configuration, e.g., a rectangle is a possibly short return time to the starting point of the work of a cluster and to the transportation facilities. A disadvantage is a higher number of the plots close each other than in a non-closing configuration, e.g., in a line or in an L-shape cluster.

In NAFORMA, there will separate personnel for transportation wherefore this is not an issue in selecting the cluster shape. Comparing an L-shape and a line-shape cluster, an advantage is the L-shape is that it is not as sensitive as a line-shape for possible systematically oriented features in land cover and land use structures than a line-shape.

An L-shape cluster was thus selected in a collaboration with the local experts (Fig. 14). The decision concerning the number of the plots remained for the sampling simulation studies, as well as the distances between the clusters.

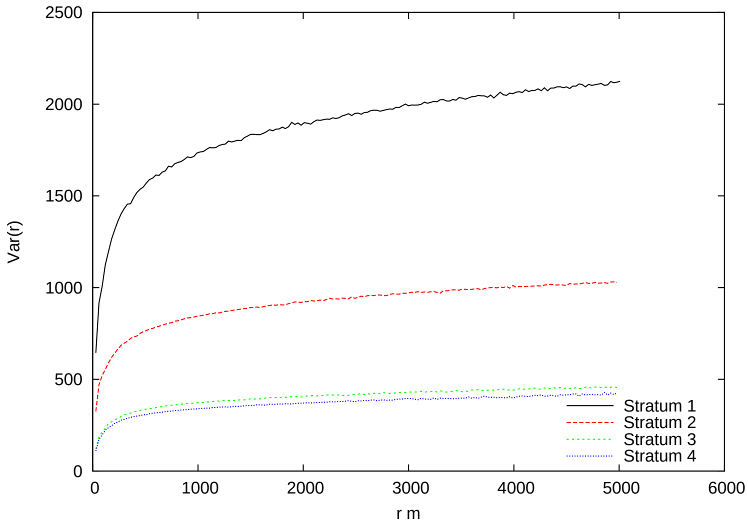

Figure 12: Semivariances of predicted volume in categories 1-4 of the re-classified vegetation types (Stratum) of Hunting map for distances 1-5000 metres.

9 Results

9.1 Double sampling for stratification

After some preliminary tests, the selected statistical framework was Double sampling for stratification. A further stratification was done on the basis of the slope variation. In double sampling for stratification, a large sample, denoted by s_a, of size n_{s_a} is drawn according to a given design. For each element, some information is either observed or predicted which information allows stratification. The information is used to

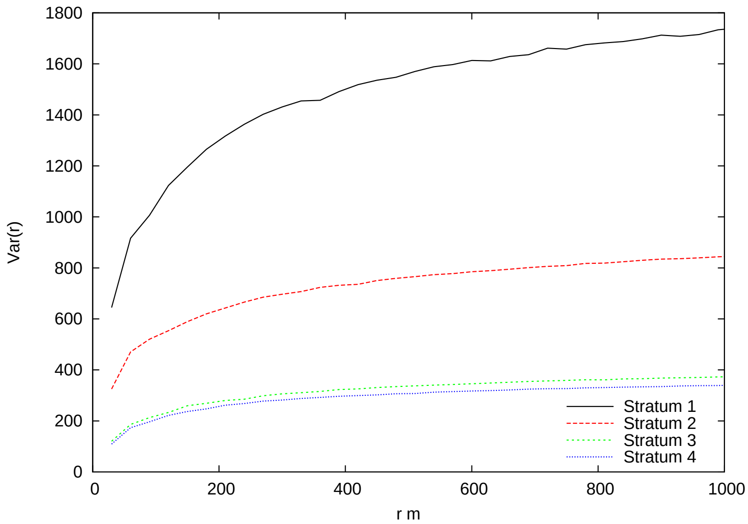

Figure 13: Semivariances of predicted volume in categories 1-4 of the re-classified vegetation types (Stratum) of Hunting map for distances 1-1000 metres.

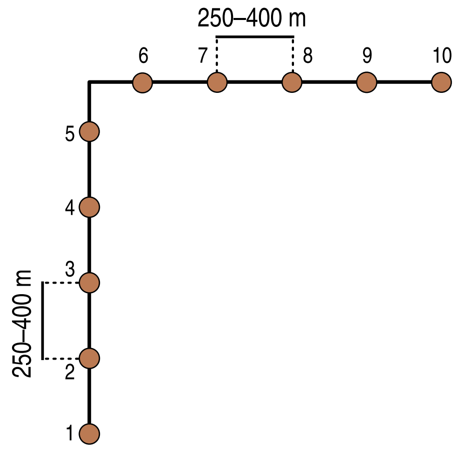

Figure 14: A field plot cluster, with plot distances to be tested.

stratify s_a into H_{s_a} disjoint strata s_{ah}, h=1, ..., H_{s_a} with n_{ah} elements in stratum h and thus the union of s_{ah},

$$s_a = \cup_{h=1}^{H_{s_a}} s_{ah} \tag{10}$$

and

$$n_{s_a} = \sum_{h=1}^{H_{s_a}} n_{ah} \tag{11}$$

From each stratum h, a sample $s_h \subset s_{ah}$ of a size of n_h is drawn and the variables of interest are measured (Cochran 1977, Särndal et al. 1992). For the total subsample s, the decomposition is

$$s = \cup_{h=1}^{H_{s_a}} s_h \tag{12}$$

and

$$n_s = \sum_{h=1}^{H_{s_a}} n_h \tag{13}$$

9.2 Application of the Double sampling for stratification

In our application, the first phase sample consisted of a dense grid of field plot clusters. A dense grid of clusters were overlaid over Tanzania using equal distances of 5 km x 5 km between the clusters (Fig. 15).

The variables used for stratification were 1) predicted average volume of growing stock for the plots of a clusters using model (6) and the estimated time to measure the plots of a cluster, the time costs including all components listed in Chapter 7. Furthermore the slope variation was used as an additional stratification criteria.

Cluster level mean volumes were calculated as the averages 1) per land area 2) per Wooded land and 3) per

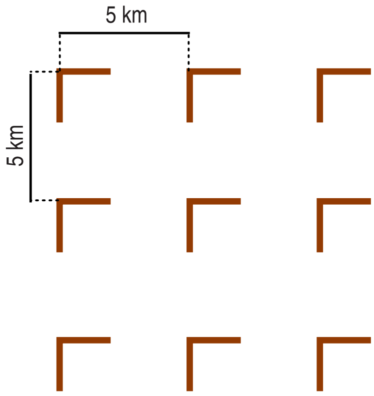

Figure 15: Field plot clusters for the first phase sample.

Forest land when using in 2) and 3) the re-classified Hunting map classes (cf. Table 2). The mean volume is simply

$$\tilde{v}_{h,l} = \frac{\sum_{i=1}^{n_{ah,l}} v_{h,i}}{n_{ah,l}} \tag{14}$$

where $v_{h,i}$ is the volume on plot i in the second phase sample in stratum h
and $n_{ah,l}$ is the number of the second phase plots on land in stratum h. The mean volumes per Wooded land and per Forest land are calculated in a similar manner just taking the summing up the respective volumes and dividing by the number of respective plots. Note that each plots within a stratum has a same areal weight.

The clusters were classified into classes for the second phase sample. Several class numbers and class interval were tested. In the selected classification, 4 volume, 3 cost and 3 slope classes were used (Table 4). The volume intervals were determined using 'optimal classification' by Neyman, see Cochran (1977, Section 5.A.7). The three cost (time) classes were used, 1) not more than one day, 2) two days and 3) more than two days per cluster.

The relative sampling intensities in different strata were selected using optimal allocation, see Cochran (1977). In our case, the sampling intensities were proportional to the the quantity

$$\hat{\sigma}_h^t / \sqrt{c_h} \tag{15}$$

where
$\hat{\sigma}_h$ is within stratum standard deviation of the mean volume of the growing stock on land on a cluster on stratum h
c is the average costs (measurement time) of a cluster within a stratum
t an exponent to be determined to control the effect of the $\hat{\sigma}_h$ on the strata weights (intensities). The exponent t was used to mitigate the effect of volume variation due to the fact that the volume is not the only interesting variable of the inventory. The sampling densities are given in Table 4.

The comparisons of the different designs in terms of costs and errors, as well as the standard errors for the used parameters of the final design, were calculated using sampling simulation and cost assessments.

The simulation was done for all designs as follows. The starting point of a grid was selected randomly within a square corresponding to the distances in south-north and west-east directions of two adjacent clusters. For a design, 1000 samples were selected. The estimate of the parameter of interest, e.g., mean volume or total volume, were computed from each sample. The mean of the estimates of the parameter values as well the standard deviation over samples were also calculated,

$$\mathrm{sd} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} \tag{16}$$

where y_i is the estimate of the parameter in i^{th} simulation and $\bar{y}$ the average of the estimates over the simulations.

On the basis of sampling error definition, the standard deviation (Eq. 16) can be used to approximate a sampling error.

The densities were adjusted to different total cost levels, and are presented in Table 5 for the costs of 1, 2.5 and 4 million US dollars for entire Tanzania and in Table 6 for Singida District. Other sampling intensities, based on the other budget constraints, can easily be calculated.

The area estimates of the classes Wooded land and Forest land (Chapter 7) for the re-classified Hunting map classes are 77.4 and 49.8 million hectares, and the estimates of corresponding total volumes for same areas 4 and 3 billion m^3. The similar area and volume estimates for Singida District are 1.9 and 1.2 million hectares and 102 and 74 million m^3 respectively.

Note that the final estimates of NAFORMA after field measurements will be based on measured variables from the second phase sample and total land area estimate of each stratum based on the first phase sample. The efficiency of the stratification depends on many factors, and also on the prediction errors. Unbiased estimates are, however, obtained despite the prediction errors. One should also note that the plots in one cluster belong always to one stratum due the definitions of the strata in question.

The cluster sizes and the rough land area estimates by strata are
strata 1-12, 10 plots, land area 83 mill. ha
strata 13-16, 8 plots, land area 4.6 mill. ha
strata 17-18 6 plots, land area 0.5 mill. ha.

Table 4: The stratification used for first phase clusters, the number of clusters in the 1st phase sample and the sampling densities ('thinning') used in the second phase.

Stratum	Measurement time of a cluster min	Mean volume on land m^3/ha	Median slope of plots °	1st phase clusters	Sampling density for 2nd phase
1.	0-<480	0-27	0-10	3080	12
2.	0-<480	27<-61	0-10	626	10
3.	0-<480	61<-118	0-10	254	8
4.	0-<480	>118	0-10	83	2
5.	480-<960	0-27	0-10	8852	13
6.	480-<960	27<-61	0-10	7282	12
7.	480-<960	61<-118	0-10	4149	9
8.	480-<960	>118	0-10	896	4
9.	960-	0-27	0-10	2252	20
10.	960-	27<-61	0-10	2766	17
11.	960-	61<-118	0-10	2033	13
12.	960-	>118	0-10	673	5
13.	0-<960	0-61	10<-20	741	7
14.	0-<960	>61	10<-20	738	4
15.	960-	0-61	10<-20	165	13
16.	960-	>61	10<-20	598	5
17.	0-	0-118	>20	243	6
18.	0-	>118	>20	94	4

The phases in sampling simulation can be summarised as follows.

- A dense grid of clusters was overlaid over Tanzania using equal distances of 5 km x 5 km between the clusters
- Cluster level mean volumes were calculated per land, as well as per classes re-classified hunting 1-6 ('wooded land') and for re-classified hunting classes 1-3 ('forest land')
- Cluster level costs (times) were calculated
- The clusters were classified into classes for the second phase sample
  - In the selected classification, 4 volume classes and 3 cost classes were used
    - The volume intervals were determined using 'optimal classification' by Neyman, see Cochran (1977)
- The sampling intensities in different strata were selected using optimal allocation, see Cochran (1977)
  - The sampling intensities are proportional to the the quantity $s^t/\sqrt{c}$
    where
    s is within stratum standard deviation of the mean volume of the growing stock on land on a cluster
    c is the average costs (measurement) time of a cluster
    t an exponent to be determined to control the effect of the s on the strata weights (intensities)
- The densities were adjusted to different total cost levels, and are presented here for 1, 2.5 and 4 million US dollars
- The standard error estimates were calculated for each design for the entire country, for strata and for the example district repeating the procedure 1000 times and taking the between sample standard deviation of the parameter estimates of interest, e.g, mean volume.

The number of the plots on land, on wooded land and on 'forest land', as well as the total costs and coefficients of variation (CV) (100*error/estimate) for four different designs for entire Tanzania are given in Table 5 and the similar figures for Singida District in Table 6. The figures for Singida District correspond to the stratified designs with the costs of 1, 2.5 and 4 million USD as well as non-stratified design of 2.6 million USD for entire Tanzania.

The variation of coefficient (CV) (relative standard error) for the area estimate of the category forest land for entire Tanzania with field measurement budget of 2.5 million US dollars is 1.2 % (Table 5). The similar figures for the budgets of 1 and 4 million US dollars are 1.9 % and 0.9 %. The CVs for the mean volume for forest land for the budget of 2.5 million US dollars is 0.85 % and that for the total volume 1.1 %. Stratification decreases particularly the errors of the volume estimates (Table 5).

One should note that the land category wooded land comprises almost total land area wherefore it is more relevant to compare the errors (CVs) of the area estimates for the category forest land.

The errors for small areas, e.g., districts, are naturally higher than for the entire country (Table 6). The final standard errors and CVs will be smaller than those in Table 6 when satellite images will be used for enhancing the estimates (Chapter 10).

Table 5: The number of the plots and the coefficient of variation for the stratified designs corresponding the costs of, 1, 2.5 and 4 Mill. USD and non-stratified designs of 2.6 Mill. USD for entire Tanzania.

	1 Mill. USD	2.5 Mill. USD	4 Mill. USD	2.6 Mill. USD No stratif.
Plots on land	13 011	32 551	52 536	35 214
Plots on wooded land	11 635	29 086	47 133	30 913
Plots on forest land	7 806	19 472	31 704	19 827
Crew days	2 517	6 259	10 189	6 600
Costs (USD)	1,006,648	2,503,600	4,075,421	2,640,009
CV				
-Area wooded land	0.77	0.44	0.33	0.39
-Area forest land	1.88	1.16	0.81	0.87
-Mean vol wooded land	0.99	0.60	0.48	1.00
-Mean vol forest land	1.54	0.85	0.69	1.14
-Total vol wooded land	0.81	0.53	0.42	1.09
-Total vol forest land	1.81	1.12	0.86	1.36

Table 6: The number of the plots and the coefficient of variation for Singida District for the stratified designs corresponding the costs of, 1, 2.5 and 4 Mill. USD and non-stratified designs of 2.6 Mill. USD for entire Tanzania.

	1 Mill. USD	2.5 Mill. USD	4 Mill. USD	2.6 Mill. USD No stratif.
Plots on land	228	544	887	500
Plots on wooded land	204	484	795	427
Plots on forest land	139	334	561	271
Crew days	50	107	169	91
Costs (USD)	19,928	42,677	67,630	36,461
CV				
-Area wooded land	8.61	4.33	3.29	3.14
-Area forest land	17.86	9.78	7.80	10.42
-Mean vol wooded land	7.88	4.75	3.86	9.82
-Mean vol forest land	13.65	6.94	5.47	8.63
-Total vol wooded land	8.32	4.33	3.87	11.07
-Total vol forest land	15.08	9.78	5.92	14.59

The locations of the clusters and field plots corresponding double sampling for stratification and a budget of 2.5 million US Dollars are shown for entire Tanzania in Fig. 16 and for Singida District in Fig. 17.

10 Recommendation for multi-source inventory for Tanzania

In the discussions with the local experts and data users, as well as in the Workshop held in July and November-December 2009, a high demand for District level forest resource information became very clear (Tomppo et al. 2009c). One reason is the fact that many important decisions are made at District level. This information need was one of the key driving forces behind the new design.

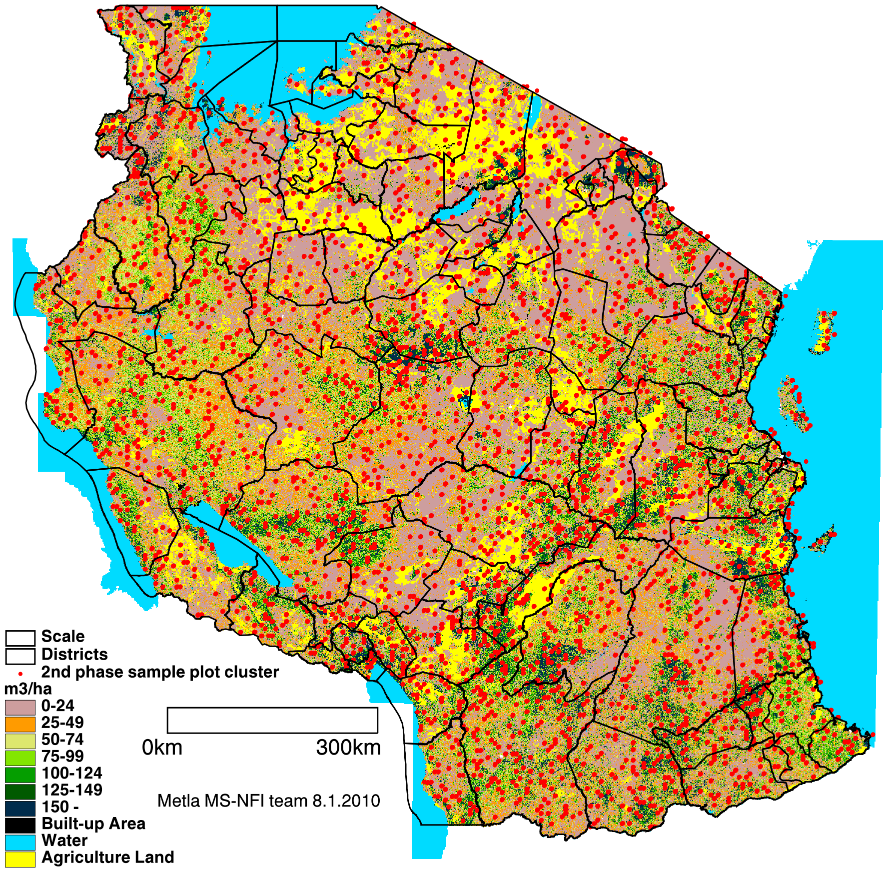

Figure 16: The location of the clusters and plots using double sampling for stratification, 32 551 plots on land.

Despite the design has been tailored for Tanzania conditions, there is still need to use ancillary data for getting applicable forest resource information for District level in case of small Districts.

The model of multi-source forest inventory (MS-NFI) used by some countries are directly applicable (e.g. Tomppo et al. 2008a and 2008b). The main purpose of the multi-source inventory is to be able to get forest resource estimates for small areas (Districts, sub-Districts). A bi-product is covering wall-to-wall thematic map form predictions of some key forest and land use variables for several forestry, economical and ecological purposes. A method that allows a simultaneous estimation of all the numerous parameters of the NAFORMA inventory is highly recommended.

The country can be covered with medium resolution data (about 20-30 metres pixel size). A feasible date of the material is year 2011 in case the field measurements will take place in 2010-2012. Optical area data have turned out to work better than microwave data in several tests, and particularly when taking into account the fact the sensor and data should be capable to distinguish several forest attributes. The image acquisition time should be in optimal case be such that the photo-synthesis activity is high.

In the operative applications of the multi-source inventories, the images from different dates are processed separately reducing a need for and atmospheric correction. When images are processed separately, it is

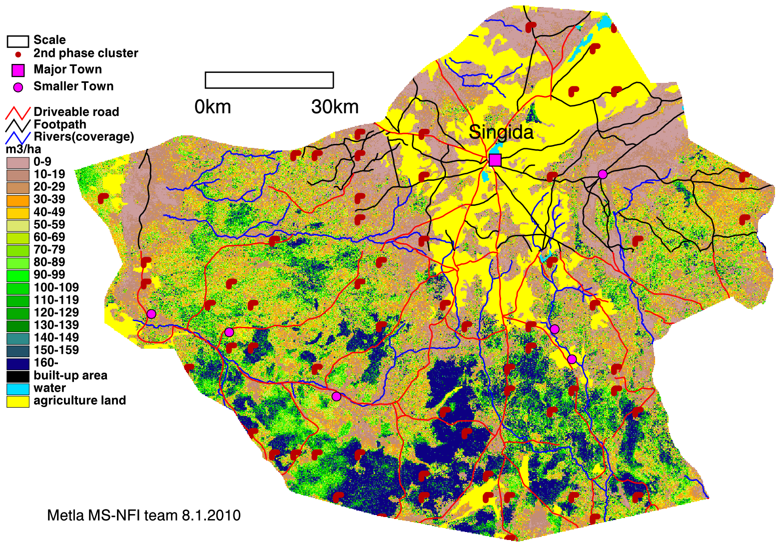

Figure 17: The location of the clusters and plots using double sampling for stratification, Singida District.

essential that the mapping between forest variables and image variables within the image is as uniform as possible. Physical correctness is, at most, a secondary issue. Atmospheric correction methods can be used within single images to enhance uniformity. One example is the haze correction algorithm is used in the Swedish k-nn product (Hagner and Olsson 2004, Tomppo 2008b).

If mosaics of several images are used for prediction of forest variables, some method of radiometric unification or normalisation has to be used between the images. The available methods are not perfect and this means that the prediction results from mosaics are not as good as with single images, provided that other factors do not affect the results. One of the most important of these factors is number of field plots within each image area. If number of field plots within an image is too small, using mosaics may yield better results even when the radiometric unification is not perfect.

The tradeoffs should be considered case-by-case when the imagery and field data are available.

An approach used in Finnish NFI and in Swedish k-nn product (Tomppo et al. 2008b), as well as in the connection of United States of Forest inventory and analysis Programme (FIA) allows to estimate all parameters simultaneously. The employed non-parametric k-nearest neighbour approach require the selection of some estimation parameters but can be adjusted in a flexible way to many types of vegetation zones and many types of remote sensing material.

Testing of the use of air-borne laser scanner data in the connection of NAFORMA, has also been discussed and elaborated.

11 Discussions on the results

We have created a method and presented the results to create sampling design for Tanzania forest resource inventory, NAFORMA.

The key factors in evaluating the optional designs were standard errors for the selected forest and land use parameters at country level and District level as well as the total costs at country level in terms of field measurement time.

Digital land use and vegetation type information together with elevation variation and road network were employed in sampling simulation, as well as the predicted volume of growing stock.

The satellite image data used for volume prediction was from 2000. The quality of that data set was better than that of the other options available. To our understanding, the changes in volume and land use since 2000 do not essentially affect the optimality of the sampling design.

The key elements of the design are:

- Double sampling for stratification
- The strata are 'imaginary', the areas represented by the set of the clusters belonging to a same stratum.
- Predictions of the volume of growing stock were used as well as costs assessments.
- The real volumes can, and of course, will deviate from the predictions. Despite this fact, the final estimates will be unbiased.
- The current strata can also be used in the coming inventories.
- Every fourth cluster will be marked to be re-measured in the coming inventories.
- The approach taken presumes some input data, in this case volume predictions and some land use/land cover area information (Hunting maps in this case). These data sets can be replaced by other similar data sets in possible other applications of the method presented.
- Otherwise the procedure can be repeated in quite a straightforward way.

In the tests carried out, the stratified design improved the efficiency significantly compared to the non-stratified alternatives (Tables 5 and 6). Stratification on the basis managed versus non-managed forests was also considered. The area of non-managed forests is not significant wherefore this option is not relevant in Tanzania and was left out from the analyses.

An exact comparison of the traditional NFMA design (FAO 2008) and the design proposed had required detailed information about the proportion for large trees and small trees for cost and error assessments. A rough idea about the efficiency of the NFMA design compared to the proposed design design in another type of forests is given in Tomppo and Katila (2008).

The relative standard error of the estimate of the land category called here “forest land” with the stratified design presented and with a total field measurement budget of 2.5 million USD is 1.2 %, and the similar errors for the mean and total volumes of the growing stock 0.9 % and 1.1 % respectively.

When assessing the costs presented, one should note that the fact that some remote clusters require a high amount of time to reach them. It has been assumed in the cost calculations that each cluster will be visited separately. However, in practice, near-by remote clusters could be measured by staying in the neighbourhood of the clusters over night which fact may reduce the total costs from the ones presented.

The recommended design allows to use the NAFORMA as the basis for UNFCCC greenhouse gas reporting of and for REDD purposes. The NAFORMA field plots cover all land categories making it possible to assess the areas and area changes of all IPCC land categories as well as carbon pools and they changes related to land category areas and land area changes (IPCC 2003, 2007). The final relevance of NAFORMA for GHG reporting depends on the availability of the data and/or models needed for carbon pool estimation, e.g., soil data and biomass models, and is thus also a matter of field measurement instructions and definitions.

The accuracy of the estimates of changes for small areas can be enhanced with additional information, e.g., remote sensing based information. A sound statistical approach is highly recommended also when using remote sensing data as an additional information.

Compatibility of the carbon pool estimates with the international definitions, e.g, with the IPCC definitions, depends on of the definitions of variables in the field measurements. To our understanding, the FAO FRA definitions have been employed everywhere when relevant making the estimates comparable with the harmonised ones (Tomppo et al. 2010).

The use of the NAFORMA for REDD purposes requires accurately geo-referenced information and small area estimates, e.g,. estimates for the areal units employed in forest management. These facts were the key driving principles in planning the current design. Although the use of the available field measurement resources has been optimized, there is still likely a need to enhance the the small area estimates using a multi-source approach.

References

Aitken R. 1977. Wilderness areas in Scotland. Unpublished Ph.D. thesis. University of Aberdeen.

Cochran, W. G. 1977. Sampling techniques. 3rd ed. New York, Wiley

Cressie, A.C. 1993. Statistics for Spatial Data. Wiley. 900 p. New York, Chichester, Totonto, Brisbane, Singapore.

FAO. 2004. National Forest Inventory Field Manual Template. Rome, August 2004. ftp://ftp.fao.org/docrep/fao/008/ae5

FAO. 2005 Global Forest Resources Assessment 2005. Progress towards sustainable forest management. FAO Forestry Paper 147, Rome, 323 p 16 E. Tomppo et al.

FAO. 2008. National Forest Monitoring and Assessment - NFMA.
http://www.fao.org/forestry/47655/en/

Gregoire, T. G. & Valentine, H.T. 2008. Sampling strategies for natural resources and the environment. Chapman & Hall/CRC, Applied Environmental Statistics. Boca Raton. 474 p. ISBN 978-1-58488-370-8.

Hagner, O., & Olsson, H. 2004. Normalisation of within-scene optical depth levels in multispectral satellite imagery using national forest inventory plot data. Proceedings from the 24th EARSeL symposium, workshop on “Remote sensing of land use and land cover”, Dubrovnik, Croatia, May 28–29, 2004.

IPCC. 2003. Penman J, Gytarsky M, Hiraishi T, Krug T, Kruger D, Pipatti R, Buendia L, Miwa K, Ngara T, Tanabe K, Wagner F (eds). Good practice guidance for land use, land-use change and forestry. Institute for Global Environmental Strategies (IGES), Hayama, Japan

IPCC. 2007. IPCC Fourth Assessment Report: Climate Change 2007. http://www.ipcc.ch/ipccreports/assessments-reports.htm

Katila, M. 2006. Empirical errors of small area estimates from the multisource National Forest Inventory in Eastern Finland. Silva Fennica 40(4): 729-742.

Langmuir E. 1984. Mountaincraft and leadership. The Scottish Sports Council/MLTB. Cordee, Leicester.

Saket, M., Branthomme, A. and Piazza, M. 2010. FAO NFMA – Support to Developing Countries on National Forest Monitoring and Assessment. In: Tomppo, E., Gschwantner, Th., Lawrence, M. & McRoberts, R.E. (eds.). National Forest Inventories - Pathways for common reporting. Springer, p. 583-594. ISBN 978-90-481-3232-4

Särndal, C.-E., Swensson, B., & Wretman, J. 1992. Model assisted survey sampling. Springer-Verlag, Inc. New York. 694 p.

Tomppo, E. & Halme, M. 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote Sensing of Environment 92: 1-20.

Tomppo, E. 2006. The Finnish National Forest Inventory. In: Kangas, A. & Maltamo, M. (eds.). Forest inventory. Methodology and applications. Managing Forest Ecosystems. Vol 10. Springer, Dordrecht. p. 179-194.

Tomppo, E., Haakana, M., Katila, M. & Peräsaari, J. 2008a. Multi-source national forest inventory - Methods and applications. Managing Forest Ecosystems 18. Springer. 374 p.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O. & Katila, M. 2008b. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment 112: 1982-1999.

Tomppo, E. & Andersson, K. (eds.) 2008. Technical review of FAO’s approach and methods for National Forest Monitoring and Assessment (NFMA). FAO, NFMA Working Paper 38. 86 p.

Tomppo, E. & Katila, M. 2008. Comparing alternative sampling designs for national and regional forest monitoring. In: Tomppo, E. & Andersson, K. (eds.). Technical review of FAO’s approach and methods for National Forest Monitoring and Assessment (NFMA). FAO, NFMA Working Paper 38: 55-86.

Tomppo, E.O., Gagliano, C., De Natale, F., Katila, M. & McRoberts, R.E. 2009a. Predicting categorical forest variables using an improved k-Nearest Neighbour estimation and Landsat imagery. Remote Sensing of Environment 113: 500-517.

Tomppo, E., Haakana, M., Katila, M., Mäkisara, K. & Peräsaari, J. 2009b. The multi-source National Forest Inventory of Finland - methods and results 2005. Working Papers of the Finnish Forest Research Institute 111. 277 p.

Tomppo, E., Malimbwi, R., Katila, M., Chamuya, R., Mäkisara, K., Peräsaari, J., Otieno, J., Nssoko, E., & Leppänen, M. 2009c. Sampling design for NAFORMA - Forest Monitoring and Assessment of Tanzania. Presentation. Workshop on NAFORMA Sampling Design, 30 November- 1 December 2009, Peacock Hotel Millenium Towers, Kijitonyama, Dar es Salaam, Tanzania.

Tomppo, E., Gschwantner, Th., Lawrence, M., McRoberts, R.E. (Eds.) 2010. National Forest Inventories Pathways for Common Reporting. 612 p. Springer. Heidelberg Dordrecht London New York. ISBN 978-90-481-3232-4.

UN-ECE/FAO. 2000. Forest Resources of Europe,CIS, North America, Australia, Japan and New Zealand. United Nations, New York and Geneva, 2000.

UNFCCC. 2008. Report of the Conference of the Parties on its thirteenth session, held in Bali from 3 to 15 December 2007. Addendum. http://unfccc.int/resource/docs/2007/cop13/eng/06a01.pdf

UNFCCC. 2009 United Nations Framework Convention on Climate Change. http://unfccc.int/2860.php