Sampling design for NAFORMA, Tanzania.

Mr. Mikko Leppanen
FAO

Sampling design for NAFORMA

Erkki Tomppo[1], Rogers Malimbwi[3], Matti Katila[1], Nurudin Chamuya[2], Kai Mäkisara[1], Jouni Peräsaari[1], Jared Otieno[2], Edwin Gerold Nssoko[2], Mikko Leppänen[4]
[1] Finnish Forest Research Institute, Vantaa, Finland
[2] Forest and Beekeeping Division of Ministry of Natural Resources
[3] Sokoine University of Agriculture
[4] Food and Agriculture Organization of the United Nations

IPCC Expert Meeting on National GHG Inventories - a Stock Taking
23-25 February, 2010, Yokohama, Japan

Background

- FAO / NFMA team has assisted countries in establishing and maintaining forest monitoring and assessment systems, 23 countries in the program world-wide
- REDD/MRV and GHG reporting needs have become an important component of forest inventories
- NAFOMRA, The National Forest Monitoring and Assessment is the first comprehensive nationwide forest inventory for Tanzania
- A tailored approach was decided to be used instead of the traditional FAO/NFMA design
- The purpose is to fulfill both forestry, REDD/MRV and international reporting needs with reasonable costs

Outline of the presentation

- Input data sets
- Methods
  - Cost estimation
  - Error estimation
- Results
- Multi-source inventory for Tanzania
- Take-Home messages

Outputs of the sampling study

- Comparisons of alternative sampling designs in terms of the errors and costs
  - Location of the NAFORMA field plots on a digital map
  - Recommendations for Multi-source inventory

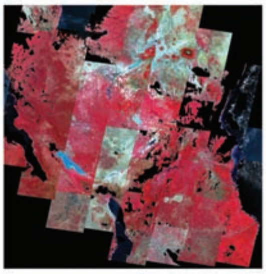
Input data sets for the sampling study

- Satellite image mosaic over Tanzania
- Hunting map over Tanzania, vegetetation types, roads, etc.
- DEM
- 11 District data from Tanzania, agregated data
- District boundaries
- Field plot data from Finland
  - Wall-to-wall volume predictions to Tanzania

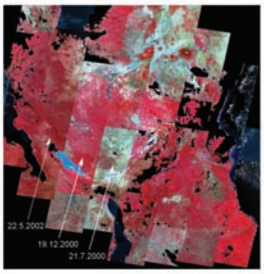
Landsat image mosaic for Tanzania

- Based on the GLS 2000 (Global Land Survey) data set from USGS
  - GLS 2000 was chosen instead of GLS 2005 because
    - Better image quality (less clouds, ETM+ sensor)
    - Closer to GIS data dates
    - The changes after 2000 were considered less significant to sampling design than the advantages above
  - 59 images Landsat 7 ETM+ images covering Tanzania
    - Downloaded from USGS to get all spectral channels
  - Three hazy images substituted with better alternatives
  - Includes images from nearly all seasons from 1999 to 2002
  - Images transformed to UTM 36 South projection with WGS84 datum
  - Coarse cloud and cloud shadow mask made manually for each image

Mosaic of Landsat Images (Top Of Atmosphere)



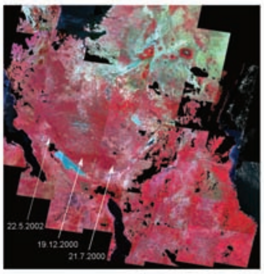
Mosaic of Landsat Images (Top Of Atmosphere)



Atmospheric correction of the images

- **MODIS MOD 09 product used as reference**
  - Surface reflectance computed from images during eight days
  - MODIS Aqua composite from 26.2.-5.3.2003 was used
    - The least cloudy alternative
  - Similar correction was done for the Finnish images using MODIS Aqua composite from 4.7.-8.7.2002
  - A transformation from the digital numbers of each Landsat image to the MODIS image was determined for Landsat channels 1 - 6
    - The mean and standard deviation were matched taking into account the different resolutions of the materials
  - The correction does force the image spectral measurements from the different imaging conditions and phenological state to same numeric scale
    - Not physically correct but usable
- **The final mosaic size 41800 columns by 43044 rows (21 gigabyte image)**

Mosaic of Atmospherically Corrected Landsat Images



Predicting volume of growing stock

Non-linear robust model for volume

- **The parameters were estimated using Finnish data, top of atmosphere reflectance, with atmosphere calibration, and non-linear estimation**
- **The final model after the calibration with the 11 District data is, see the next slide**

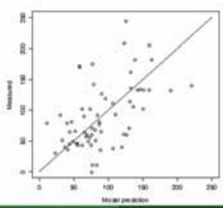
$$vol = c * exp(a + b_1 * band_3/band_2 + b_2 * band_7/band_5) + \epsilon$$

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

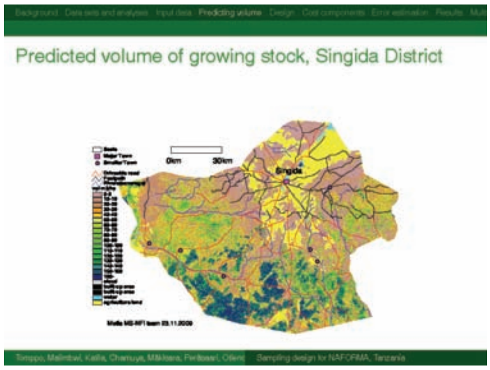
- **The model explained 75 % of the volume variation**
- **Other variables, such as brightness, greenness, wetness, were also tested**

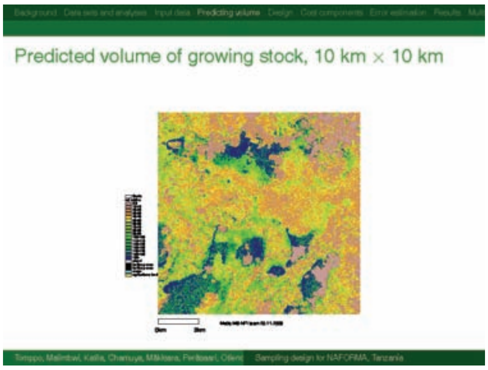
Volume calibration

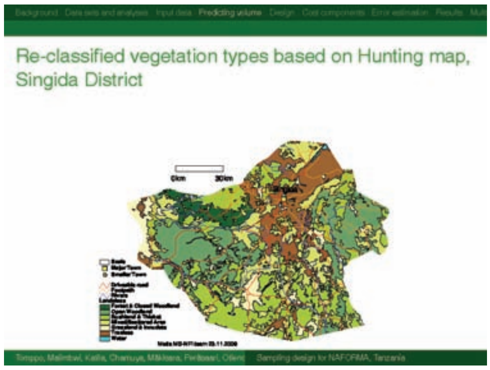
- **The volumes were calibrated using 11 District data and a linear calibration**
- **The fit after calibration**

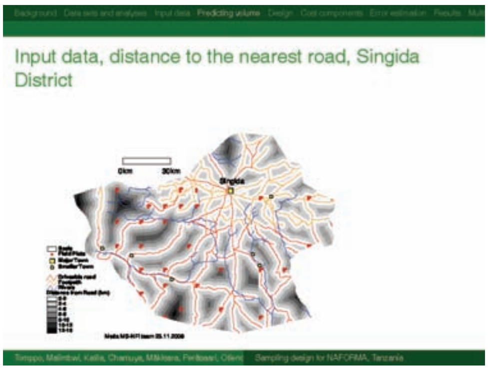


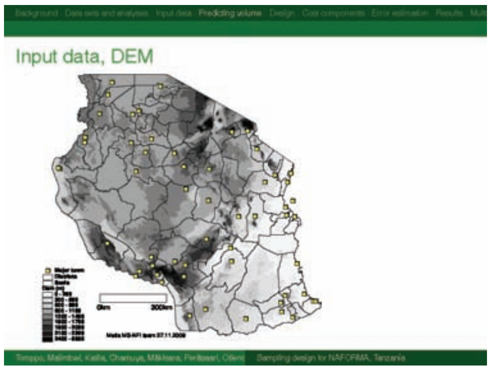

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results
Predicted volume of growing stock
0km 300km
Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Peltoasari, Otieno
Sampling design for NAFORMA, Tanzania


Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results
Predicted volume of growing stock, Singida District
0km 30km
Singida
Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Peltoasari, Otieno
Sampling design for NAFORMA, Tanzania

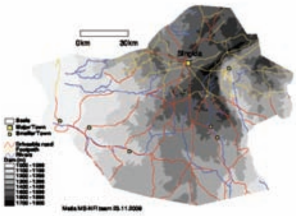

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results
Predicted volume of growing stock, 10 km × 10 km
Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Peltoasari, Otieno
Sampling design for NAFORMA, Tanzania


Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results
Re-classified vegetation types based on Hunting map, Singida District
0km 30km
Singida
Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Peltoasari, Otieno
Sampling design for NAFORMA, Tanzania


Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results
Input data, distance to the nearest road, Singida District
0km 30km
Singida
Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Peltoasari, Otieno
Sampling design for NAFORMA, Tanzania


Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results
Input data, DEM
0km 300km
Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Peltoasari, Otieno
Sampling design for NAFORMA, Tanzania

Input data, DEM, Singida District



Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Elements of a design

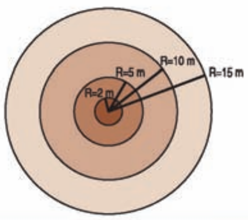
- Sample plot size and shape
- Spatial layout of the plots
  - Detached plots / plot clusters
  - Distances between the plots
  - Distances between possible clusters
- The solution is far from trivial and depends also on the parameter in question, area estimate, volume estimate, estimate of rare events
- Practical things must be taken into considerations, cost, the measurement unit should be a work-load of one day for a field crew

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perälaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Field plot

A concentric field plot in cluster designs, max radius 15 m was selected on the basis of the earlier local tests



Species name and dbh of all measured trees will be recorded in each plot in the following manner

1) Within 2 m radius: all trees with dbh > 0 cm will be recorded
2) Within 5 m radius; all trees with dbh > 5 cm will be recorded
3) Within 10 m radius; all trees with dbh > 10 cm will be recorded
4) Within 15 m radius; all trees with dbh > 20 cm will be recorded

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perälaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Variogram and semivariance, tools to assess plot distances

Variogram of a process Z, e.g., mean volume of the growing stock

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

$\gamma(x,y)$ is called semivariogram.

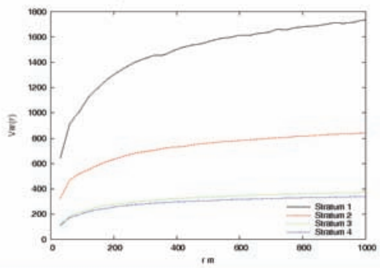
A robust estimate for semivariance

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

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perälaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Semivariances of predicted volume in strata 1-4

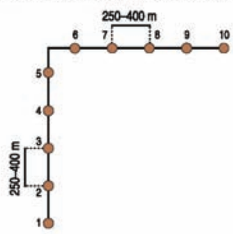


Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perälaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

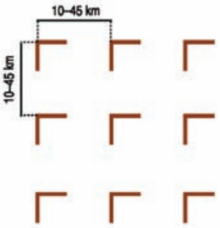
A field plot cluster, with plot distances to be tested

A distance of 250 metres was confirmed with sampling studies



Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perälaari, Otieno Sampling design for NAFORMA, Tanzania

Field plot clusters with basic starting distances



Summary of components considered in time calculations

- Driving to a cluster from the lodgement (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)

Walking distance and time, cont

Figure 3. 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 multi-source NFI.

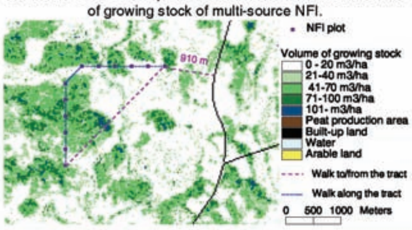


Table 4. For systematic cluster samples of L-shape, following walking speed assumptions and average plot measurement times were used, broken down to 'Hunting map' vegetation classes.

VEGCODE		Pre-strat	VEGNBR DESCRIPTION / 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

VEGCODE		Pre-strat	VEGNBR DESCRIPTION / vegtype	walk speed, min/km.	plot meas min
GSc	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

Steps taken in determining error estimates

- Select a set of potential sampling designs
- For each design, simulate a large number of samples, e.g., 1000, with different starting points on the volume and land cover map
- Calculate estimates from each sample
- Calculate the standard deviation of the estimates, it can be considered as a sampling error

The parameters for which the errors were studied

a) Area of land classes (ha), grouped Hunting map strata 1-3 ('forest') 1-6 (wooded land)
b) Mean tree stem volume of growing stock (m^3/ha), Hunting strata 1-3 and 1-6
c) Total volume of growing stock (m^3), Hunting strata 1-3 and 1-6

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Examples of possible alternative sampling designs

1. NFMA design, a NFMA tract distance in both latitude and longitude is one degree, and its desified versions
2. Systematic cluster designs with a plot distance of 250 m apart from each other (no stratification)
3. Stratified cluster designs, the clusters distances and the number of the plots per cluster vary by strata

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perässaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Double sampling for stratification

- The selected statistical framework was Double sampling for stratification, see, e.g., Cochran (1977)
- A dense grid of clusters were overlaid over Tanzania using equal distances of 5 km x 5 km between the clusters
- Cluster level mean volumes were calculated per land
- Cluster level costs (times) were calculated
- The clusters were classified into classes for the second phase sample
- The second phase is a sub-sample of the first phase sample and will be measured in the field
- Several class numbers and class intervals were tested
  - 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)

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perässaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Double sampling for stratification, cont

- The sampling intensities in different strata were selected using optimal allocation and are proportional to the the quantity (Cochran, 1977)
  -
    $$s^t/\sqrt{c}$$
    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 final estimates are based on measured variables from the second phase sample and area estimates of the strata based on the first phase sample

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perässaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Table 5. 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

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perässaari, Otieno Sampling design for NAFORMA, Tanzania

Background Data sets and analyses Input data Predicting volume Design Cost components Error estimation Results Mult

Double sampling for stratification, cont

- 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
- The error estimates and plot numbers are presented for three levels of total costs, as given above, 1, 2.5 and 4 million US Dollars
  - The rough areas of 1-6 and 1-3 classes are 77.4 and 49.8 million hectares and the volumes 4 and 3 billion m^3
- The result are presented also for Singida District
  - The rough areas of 1-6 and 1-3 classes are 1.9 and 1.2 million hectares and the volumes 102 and 74 million m^3

Tomppo, Malimbwi, Katila, Chamuya, Mäkisara, Perässaari, Otieno Sampling design for NAFORMA, Tanzania

The number of the plots on land, on wooded land, on 'forest land', total costs and variation of coefficients (cv) (100*error/estimate) for four different designs for entire Tanzania.

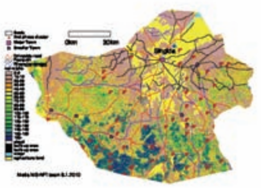
	1 Mill. USD	2.5 Mill. USD	4 Mill. USD
Plots on land	13 011	32 551	52 536
Plots on 1-6	11 635	29 086	47 133
Plots on 1-3	7 806	19 472	31 704
Crew days	2 517	6 259	10 189
Costs (USD)	1,006,648	2,503,600	4,075,421
CV			
-Area 1-6	0.77	0.44	0.33
-Area 1-3	1.88	1.16	0.81
-Mean vol 1-6	0.99	0.60	0.48
-Mean vol 1-3	1.54	0.85	0.69
-Total vol 1-6	0.81	0.53	0.42
-Total vol 1-3	1.81	1.12	0.86

The number of the plots on land, on wooded land, on 'forest land', total costs and variation of coefficients (cv) (100*error/estimate) for four different designs for Singida. The designs correspond about 2 and 2.5 mill. USD for entire Tanzania.

	1 Mill. USD	2.5 Mill. USD	4 Mill. USD
Plots on land	228	544	887
Plots on 1-6	204	484	795
Plots on 1-3	139	334	561
Crew days	50	107	169
Costs (USD)	19,928	42,677	67,630
CV			
-Area 1-6	8.61	4.33	3.29
-Area 1-3	17.86	9.78	7.80
-Mean vol 1-6	7.88	4.75	3.86
-Mean vol 1-3	13.65	6.94	5.47
-Total vol 1-6	8.32	4.33	3.87
-Total vol 1-3	15.08	9.78	5.92

The location of the plots in a stratified design, 33471 plots, about 3500 clusters

The location of the plots in a stratified design, Singida District

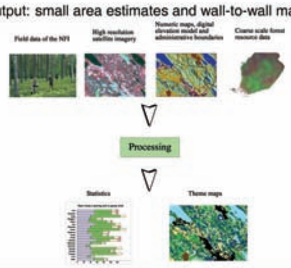


Multi-source NAFORMA for Tanzania

- The objectives
  - To be able to calculate forest resource information for smaller than what is possible using field data only, i.e., District level
    - All parameters are needed, statistics, wall-to-wall maps
  - To make the NAFORMA more cost-efficient
- The feasibility of different remote material should be thoroughly investigated keeping in mind the needs, forest management, REDD/MRV, etc.
  - In addition to medium resolution space-borne optical data RS data and LiDAR, high or very high resolution RS space-borne optical data could be competitive

The multi-source NAFORMA, cont

Output: small area estimates and wall-to-wall maps



Take-Home messages and conclusions

1. The data and methods for preparing a sampling design for Tanzania
2. Plan the forest inventory and monitoring system thoroughly, including the roles of field data and remote sensing data as well as field sampling design
3. Tailor the system for local conditions
4. Use internationally accepted, unique definitions to make the estimates comparable between countries
5. With a thorough and tailored planning, you can get much more detailed and accurate information, and with reduced costs compared to the situation without planning

Take-Home messages and conclusions, cont

6. A statistically sound method to allocate more measurements for areas with a high volume variation and low measurement cost and vice-versa
7. The combination of field data and remote sensing data is always efficient - field data ARE ALWAYS MANDATORY
8. With a thorough planning, you can, and YOU MUST GET, 'many birds by one stone', examples are
   - information for forestry
   - information for REDD/MRV purposes
   - information for international reporting
   - This is always necessary, even more when the resource are limited

A few References

- Tomppo, E., Gschwantner, Th., Lawrence, M., McRoberts, R.E. (Eds.) National Forest Inventories - Pathways for Common Reporting. 1st Edition., 2010, XXVI, 612 p. 10 illus. in color., Hardcover. Springer. ISBN: 978-90-481-3232-4
  - Includes National Forest Inventories for over 40 countries, covering more than 2.4 billion of the forests of the Globe. (All major forest countries, Brazil, Canada, China, Russia, USA, etc, plus about 30 European countries
  - The roles of NFIs in GHG reporting is described for each country
- Tomppo, E., Haakana, M., Katila, M. & Peräsaari, J. 2008. Multi-source national forest inventory - Methods and applications. Managing Forest Ecosystems 18. Springer. 374 p.

We sincerely thank the Organisers and Audience!