Designing NAFORMA - National Forest Inventory for Tanzania for Forestry and REDD+ MRV

Erkki Tomppo[1,4], Rogers Malimbwi[3,1],, Matti Katila[1], Nurudin Chamuya[2], Kai Mäkisara[1], Jouni Peräsaari[1], Jared Otieno[2], Søren Dalsgaard [2], Eliakimu Zahabu[3], 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, Rome

Regional Course on REDD+ MRV, NFI and Monitoring

11 - July 2011, Sokoine University of Agriculture, Morogoro, Tanzania

Outline of the lectures

- Tanzania sampling design, principles, methods and results
- Estimation and error estimation (results) using field data alone
-
    - preliminary results for Rufiji District
- Estimation and error estimation (results) using combined field data and remote sensing data, Multi-Source NAFORMA
    - input data sets
    - methods
    - very draft examples of the results

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
- NAFORMA, 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 **forestry**, REDD+ MRV and international reporting needs with reasonable costs

Background, cont

- Important forestry decisions are made often at small area level, e.g., in Tanzanian at District and Village level
  - High demand for small area estimates
  - Multi-source inventory plays an important role
- Another demand for the use of remote sensing arises from the large area in question and the lack of historical data

Tanzania sampling study

- Comparisons of alternative sampling designs in terms of the errors and costs
- Outputs
  - Sampling design for Tanzania fulfilling bot forestry and REDD+ MRV needs
    - Location of the NAFORMA field plots with the given budget
    - Recommendations for Multi-source inventory

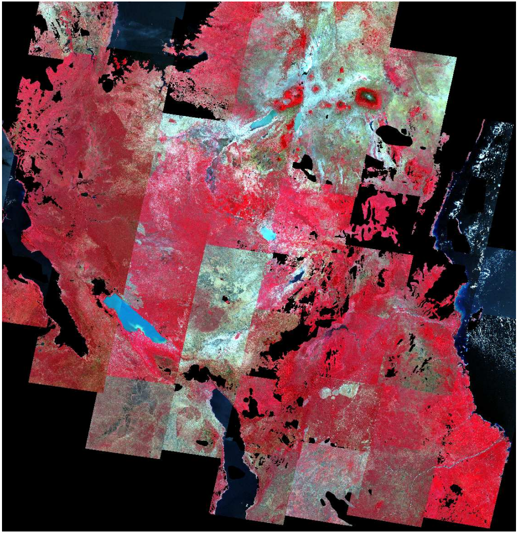
Input data sets for the sampling study

- Wall-to-wall volume predictions for entire Tanzania based on
  - Satellite image mosaic over Tanzania
  - Field plot data from Finland
  - Volume estimates from 11 District from Tanzania, aggregated data, for calibration
- Two digital maps over Tanzania, vegetation types, roads, etc., called Hunting map and Africover (FAO)
- DEM
- District boundaries

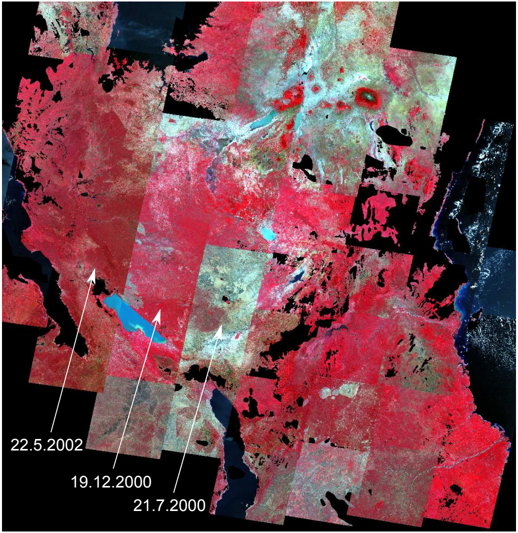
Landsat image mosaic for Tanzania

- 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)



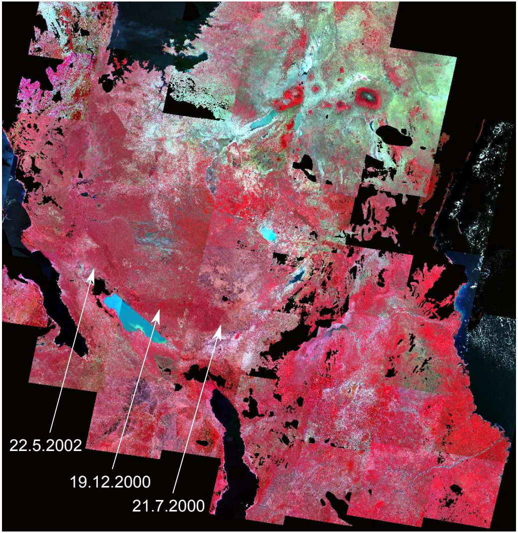
Mosaic of Landsat Images (Top Of Atmosphere)



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

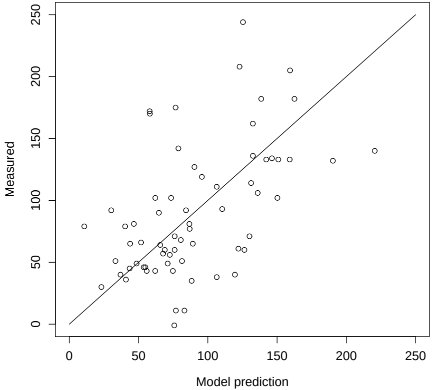
$$vol = c * exp(a + b_1 * band_3/band_2 + b_2 * band_7/band_5) + \epsilon$$

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

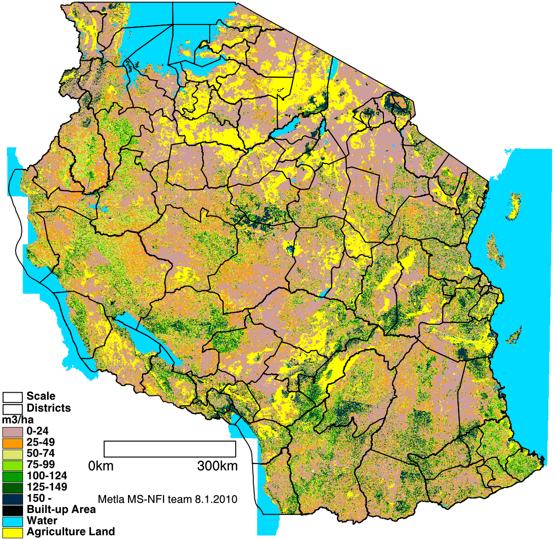
- The model explained 75 % of the volume variation
- Other variables, such as brightness, greenness, wetness, were also tested

Volume calibration

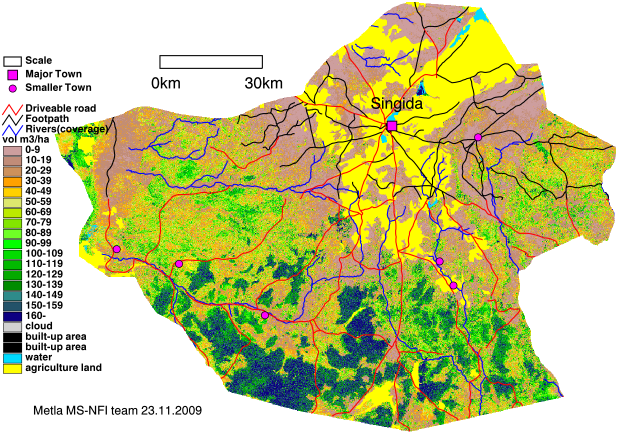
- The volumes were calibrated using 11 District data and a linear calibration
- The fit after calibration



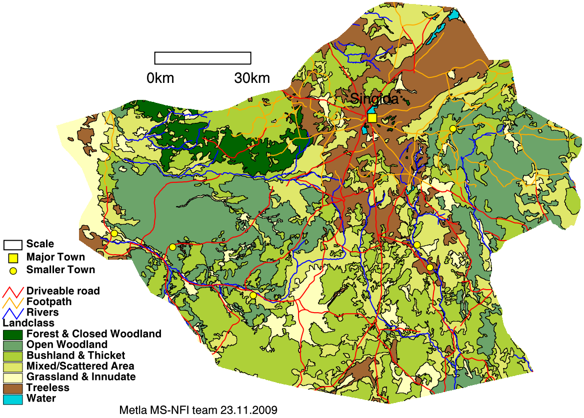
Predicted volume of growing stock



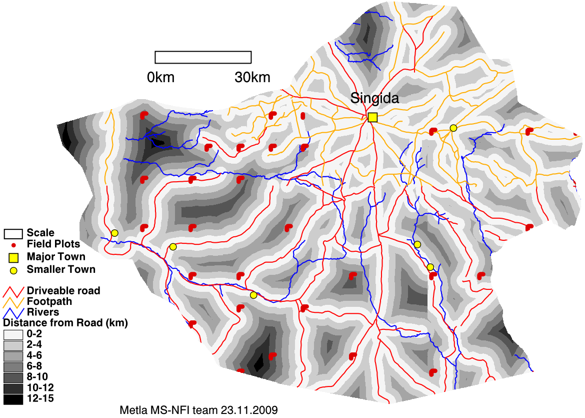
Predicted volume of growing stock, Singida District



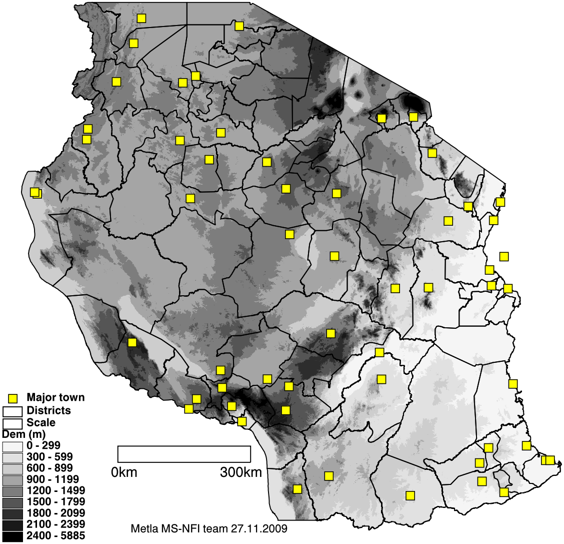
Re-classified vegetation types based on Hunting map, Singida District



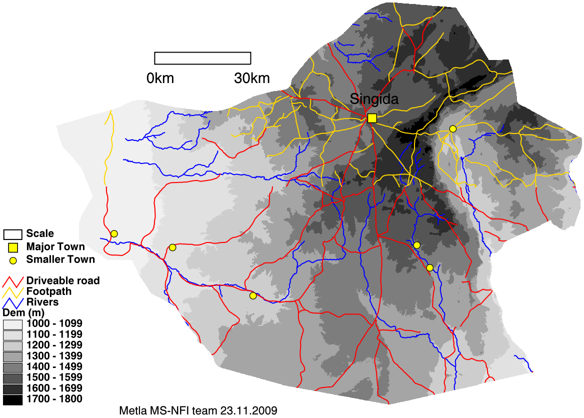
Input data, distance to the nearest road, Singida District



Input data, DEM



Input data, DEM, Singida District

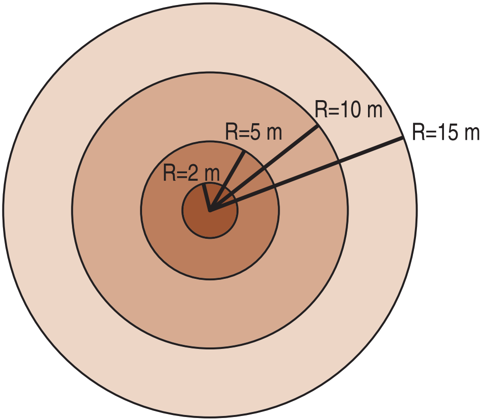


Elements of a design

- 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(s) to be estimated, area estimates, volume estimates, estimates 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

Field plot

A concentric field plot in cluster design, max radius 15 m was selected based on the earlier local studies

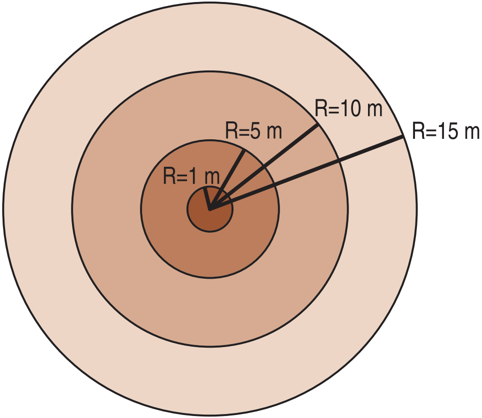


Species name and dbh of all measured trees will be recorded in each plot in the following manner

1) Within 2 m radius: all trees with dbh ≥ 1 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

Field plot since May 2011

A concentric field plot in cluster design, max radius 15 m was selected based on the earlier local studies, 2 m was reduced to 1 m and the plot measured only on the permanent plots in 2011



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

1) Within 1 m radius: all trees with dbh ≥ 1 cm will be recorded Only on permanent plots

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

Variogram and semivariance, tools to assess plot distances

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

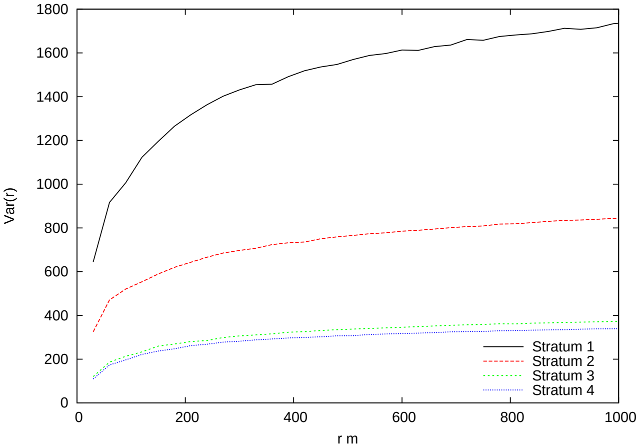
$$2\gamma(x, y) = E\left(|Z(x) - Z(y)|^2\right)$$

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

A robust estimate for semivariance

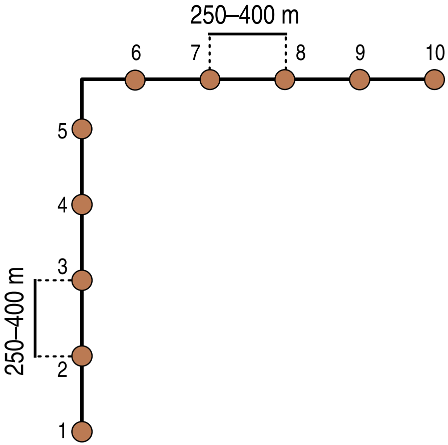
$$\hat{\gamma}(r) = \frac{(1/N(r)\sum_{S(r)}|y_i - y_j|^{1/2})^4}{0.914 + 0.988/N(r)}$$

Semivariances of the predicted volume in wooded land strata (1-4) for assessing feasible plot distances, support 200-300 m



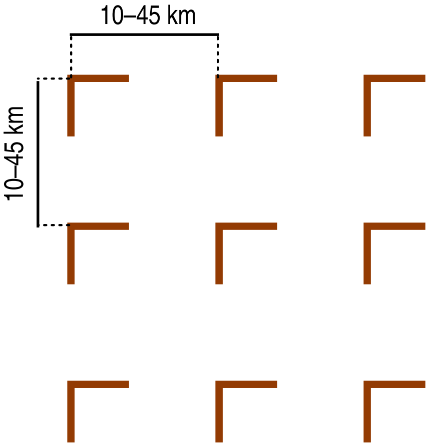
A field plot cluster, with plot distances tested

A distance of 250 metres was confirmed with sampling studies



The clusters distances in the first phase sample was 5 km $\times$ 5 km

Examples of the cluster distances for the second phase sample

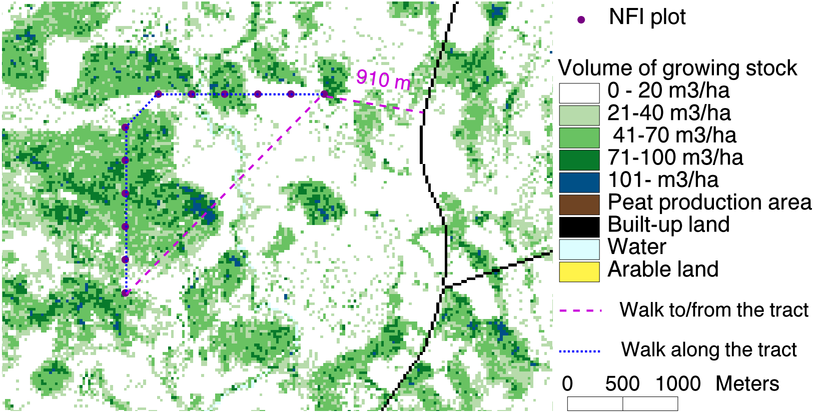


Summary of components considered in time calculations

- 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)

Walking distance and time

The minimum Euclidean distance (910 m) from the road to the closest field plot of a, cluster. The mean volume predictions are as given above.



Walking speed assumptions and average plot measurement times for systematic cluster samples of L-shape, 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

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, 3 cost classes and 3 slope classes were used
        - The volume intervals were determined using 'optimal classification' by Neyman, see Cochran (1977)

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 mitigate the effect of s on the strata weights when s is estimated using one variable only
- The sampling intensities were adjusted to different total cost levels, e.g., 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

Steps taken in assessing sampling errors

- For each design, simulate a large number, T, of samples on the volume and land cover map with different starting points of the cluster grid, e..g, 1000 samples
  - A new starting point was randomly selected for the first phase grid for the double sampling for stratification
- Calculate the estimate, $\hat{m}$ of a parameter, m, in question from each sample t
- Calculate the standard deviation of the estimates, $sd = \sqrt{\sum_{t=1}^{T} (\hat{m}_t - \bar{m})/(T-1)}$, where $\bar{m} = \sum_{i=1}^{T} \hat{m}_t/T$
  The standard deviation, sd, can be considered as a sampling error

Table 5. The stratification used for first phase clusters, the number of clusters in the 1st phase sample and the sampling intensities ('thinning') used in the second phase.

Stratum	Measurement time of a cluster min	Mean volume on land m^3/ha	Median slope of plots $^\circ$	1st phase clusters	Sampling intensity 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

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

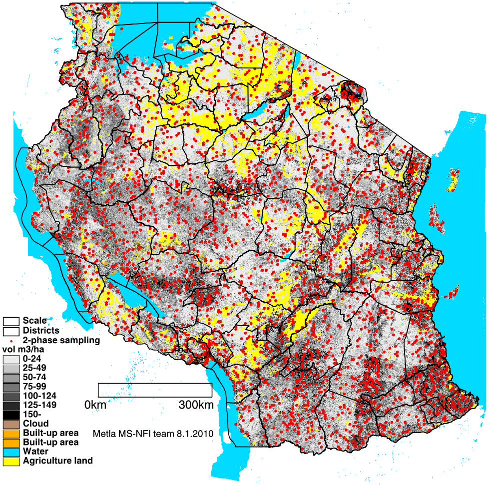
The number of the plots on land, on wooded land, on 'forest land', total costs and variation of coefficients (CV) (100*error/estimate) for four different designs for entire Tanzania.

	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

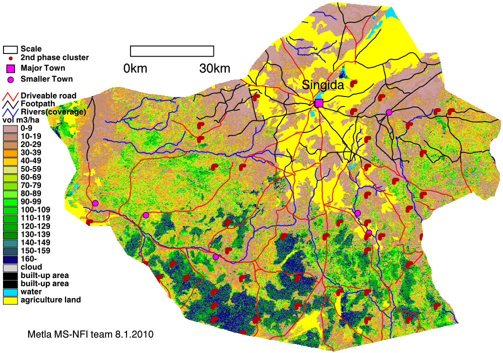
The number of the plots on land, on wooded land, on 'forest land', total costs and variation of coefficients (CV) (100*error/estimate) for four different designs for Singida. The designs correspond about 2 and 2.5 mill. USD for entire Tanzania.

	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



The role of remote sensing in NAFORMA

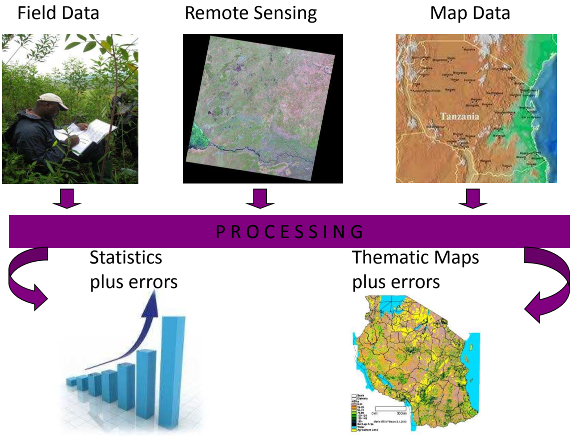
- Statistics for smaller area units than what is possible using field data only
  - District and Village level statistics for forestry and environment planning
  - Information for REDD+ MRV
- Accurately localized information, maps
  - timber procurement
  - REDD+ MRV
  - ecological and other studies

The data sets needed

1. Field data
2. Vegetation maps, LULC under process
3. Hunting technical maps (IRA)
4. Roads
5. Built-up land (a sub-category on Hunting map?)
6. DEM, resolution to be decided
7. Satellite images, (1990 ??) -2010 Landsat TM and ETM+ plus additional material
8. District and Village boundaries

-

MS-NFI


Field Data
Remote Sensing
Map Data
Tanzania
PROCESSING
Statistics
plus errors
Thematic Maps
plus errors

A non-parametric k-NN estimation with covariates

$\Longrightarrow$ Distance metric d in the *covariate space*, i.e. in the space of image features $\times$ coarse scale forest variables

$\Longrightarrow$ For the pixel p to be analysed, compute $d_{p_i,p} = \|p_i - p\|$ to each pixel p_i whose ground truth is known (to pixel with field plot i)

$\Longrightarrow$ for $k \approx$ 5-10, let $p_{(1)}, p_{(2)}, \ldots, p_{(k)}$ be the k nearest field plot pixels with respect to d

$\Longrightarrow$ The weight of field plot i to pixel p is defined as

$$w_{i,p} = \frac{1}{d^t_{p_i,p}} / \sum_{j=1}^{k} \frac{1}{d^t_{p_{(j)},p}}$$

if pixel p_i is among the k nearest to p, otherwise $w_{i,p} = 0$

k-NN estimation, cont

In the image analysis process, weights $w_{i,p}$ are summed over pixels p by computation units.
The weight of the field plot i to computation unit u, '**plot expansion factor**', is defined as

$$c_{i,u} = \sum_{p \in u} w_{i,p}$$

The estimation returns to the estimation with pure field data, e.g. mean timber volumes estimates by forestry land strata are

$$v = \frac{\sum_{i \in I_s} c_{i,u} v_{i,t}}{\sum_{i \in I_s} c_{i,u}} ,$$

where $v_{i,t}$ is the volume per hectares of the timber assortment t in sample plot i.

Background of REDD

- COP 11, 2005, Montreal,The first proposal by Papua New Guinea and Costa Rica, **RED**
- COP 13, Bali, 2007, Paragraph 1 b iii) of the Bali Action Plan: "Policy approaches and positive incentives on issues relating to reducing emissions from deforestation and forest degradation in developing countries" (**REDD**)
- COP 14, Poznan, 2008, "and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks" (**REDD+**)
- COP 16, Cancun, 2010, 70. Encourages developing country Parties to contribute to mitigation actions in the forest sector by undertaking the following activities, ...
  - (a) Reducing emissions from deforestation;
  - (b) Reducing emissions from forest degradation;
  - (c) Conservation of forest carbon stocks;

Background of REDD, cont

In the development of its work programme, the SBSTA is requested to:

- (a) Identify land use, land-use change and forestry activities in developing countries, in particular those that are linked to the drivers of deforestation and forest degradation ... associated methodological issues to estimate emissions
- (c) Develop as necessary, modalities for measuring, reporting and verifying anthropogenic forest-related emissions by sources and removals by sinks, forest carbon stocks, forest carbon stock and forest area changes resulting from the implementation of activities ... with any guidance for measuring, reporting and verification of nationally appropriate mitigation actions by developing country Parties agreed by the Conference of the Parties,

Understanding REDD

- What is measured
  - Deforestation
  - Degradation
  - Enhancement
- Reference level
  - Historical
  - Projected

NAFORMA and options for REDD+ MRV in Tanzania

- Strategic level
  - The NAFORMA permanent field plots for assessing changes in the future
    - Additional permanent plots if needed in the continuation
  - Reference level, one option is to use the NAFORMA (permanent) field plots and old remote sensing data for these plots
- Operational level
  - One option for information could be the multi-source NAFORMA

Take-Home messages and conclusions

1. Plan the forest inventory and monitoring system thoroughly, including the roles of field data and remote sensing data
2. Tailor the system for the local needs and conditions
3. REDD+ MRV requirements very demanding. Some details are still open
4. NAFORMA and MS-NAFORMA have potentials to respond to the information needs
5. NFIs have potentials to respond to the needs
6. Double sampling for stratification seems to be a relevant approach to allocate more measurements to areas with a high variation of the variables of interest and low measurement cost and vice-versa

Take-Home messages and conclusions, cont

7. Use internationally accepted, unique definitions to make the estimates comparable between countries
8. Problems or challenges with the approach used
   - needs covering predictions of the variables used, as well data to asses the costs
   - the solution depends on how 'good' the predictions are
   - different parameters may presume different sampling designs
     - this holds also for the other approaches
   - The efficiency may change over time, temporary plots are one solution to this problem
8. In estimation, the combination of field data and remote sensing data is always efficient - field data are always mandatory

A few References

- Tomppo, E., Katila, M., Mäkisara, K., Peräsaari, J., Malimbwi, R. Chamuya, N., Otieno, J. Dalsgaard, S., and Leppänen, M. 2010. 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. A report. FAO, Rome, Avalaible at: http://dl.dropbox.com/u/6700344/NAFORMA%20Sampling%20Study.pdf.
- 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
- Tomppo, E., Heikkinen, J., Henttonen, H., Ihalainen, A., Katila, M., Mäkelä, H., Tuomainen, T., Vainikainen, N. 2011. Designing and conducting a forest inventory - case: 9th National Forest Inventory of Finland. Managing Forest Ecosystems 22. Springer. 305 p. ISBN 978-94-007-1651-3.

A few References, cont

- The Little REDD+ Book. 2009. Parker, C. and Mitcell, A. Global Canopy Foundation. Oxford. UK.
- UNFCCC. 2011. Cancun Agreements: Decisions. http://unfccc.int/documentation/decisions/items/3597.php?such=j&volltext=%22
- McRoberts, R., Tomppo, E.O., Finley, A.O. & Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing of Environment 111: 466-480.
- Tomppo, E., Haakana, M., Katila, M. & Peräsaari, J. 2008. Multi-source national forest inventory - Methods and applications. Managing Forest Ecosystems 18. Springer. 374 p.
- Tomppo, E., Haakana, M., Katila, M., Mäkisara, K. & Peräsaari, J. 2009. The multi-source National Forest Inventory of Finland - methods and results 2005. Working Papers of the Finnish Forest Research Institute 111. 277 p.

We thank the Organizers and Participants!