Expert consultation on National Forest Monitoring and Assessment (NFMA): Meeting evolving Needs

26-28 November, 2008

FAO, Rome

Technical Review of FAO's Approach and Methods for National Forest Monitoring and Assessment (NFMA)

Erkki Tomppo and Krister Andersson

– The scope of the work –

- Independent scientific examination to contribute sampling design and statistical framework to estimate biophysical and socio-economic parameters
- To stimulate discussions about how NFMA can meet evolving needs and new challenges
- Sampling simulations
  - Comparisons of some alternative sampling designs
  - The error levels are same for some key parameters
  - The time consumptions are compared
- Illustrations of different types of analysis using interview data
- General comments and recommendations on sampling design and statistical analysis

Expert consultations

Review of NFMA

Simulation of sampling errors

26-28 November, 2008

FAO, Rome

Erkki Tomppo

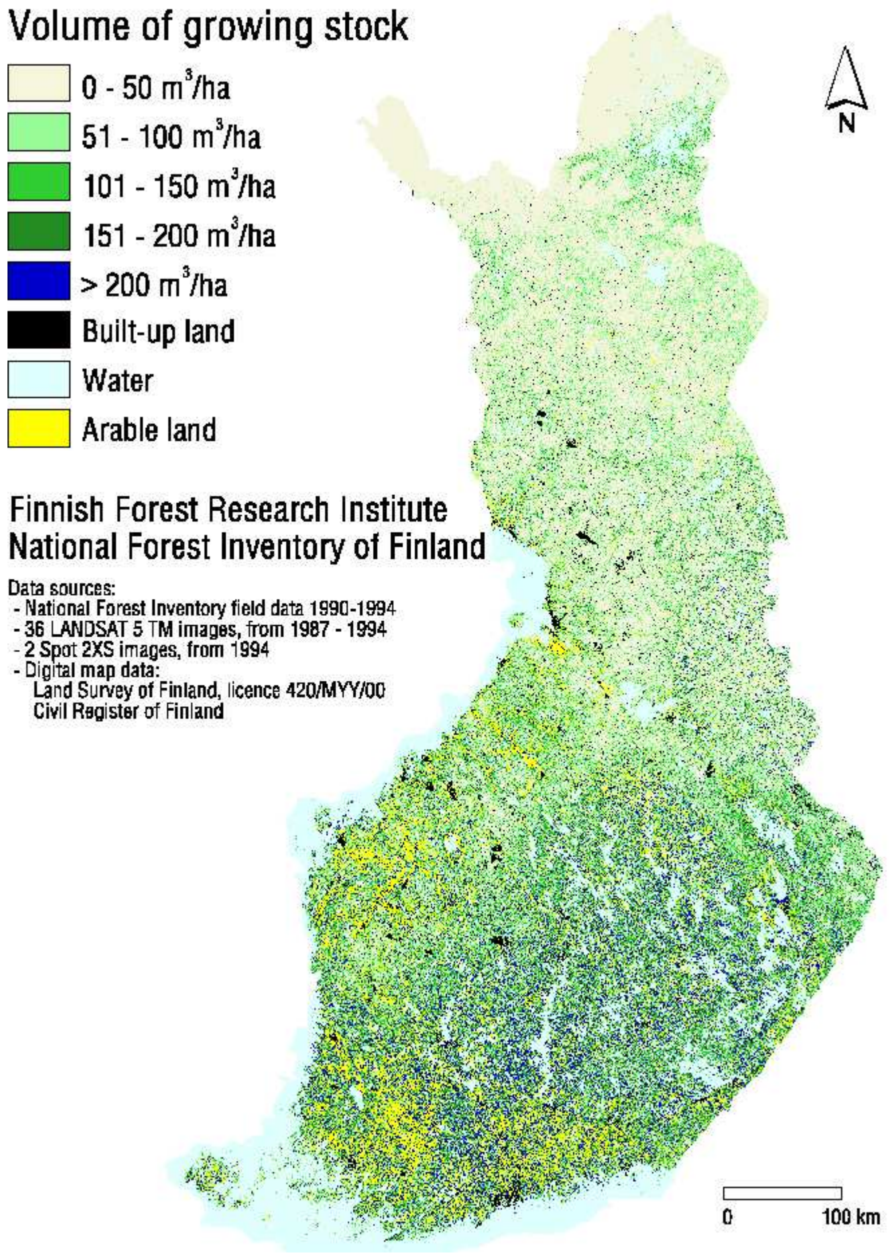
– Basic principles and facts in survey sampling –

- Each element in the population should have a positive inclusion probability
- The analysis concerns that set of the elements which have a positive inclusion probability
- An efficiency requirement: each observation should bring new information
- Accessing each unit often requires lot of time
- A key question: how to determine the spatial layout (design) of the sampling units?
  - What is 'optimal' distance between the observations
  - Possibilities to study the design and plot layout
    - A model of land use and forest structure, ' a census of a land area'
    - Semivariance (an estimator of a variogram) can be used to assess what is the optimal distance between plots and tracts

– Sampling simulation –

- A model (census) of the land area: output maps of Finnish multi-source inventory
  - Predictions of some core variables for each square of 25 m $\times$ 25 m, land class, volumes by tree species, age of trees, etc
- Two groups of the designs, sparse and dense
- The designs selected so that the error levels are same for some key parameters in each group
- The time consumptions can be compared

The volume of growing stock - Finnish multi-source NFI



– The parameters for which the errors were studied –

a) area of forestry land (ha)

b) mean tree stem volume of growing stock (m^3/ha)

c) total tree stem volume of growing stock (m^3)

d) mean volume of saw timber (m^3/ha)

e) total volume of saw timber (m^3)

f) mean tree stem volume of broad leaved trees except birch (m^3/ha)

g) total tree stem volume of broad leaved trees except birch (m^3)

- BD variables and change estimates can be added

– Steps taken in determining error estimates –

- Simulate a large number of samples, e.g., 1000, with different starting points from the thematic maps
- Calculate estimates from each sample
- Calculate the standard deviation of the estimates, it can be considered as a sampling error

– Alternative sparse designs –

1. A sparse NFMA design, a NFMA tract distance in both latitude and longitude is one degree, the current NFMA design.
2. A Sparse grid of detached plots with the intervals of 37 km $\times$ 37 km, called here a sparse Eurogrid.
3. A sparse cluster design, a cluster consisting of 12 plots located on the sides of a half rectangle (Figure 2) with a distance of 300 m apart from each other and with cluster distances of 80 km $\times$ 80 km (Cluster, no stratification).

– Alternative dense designs –

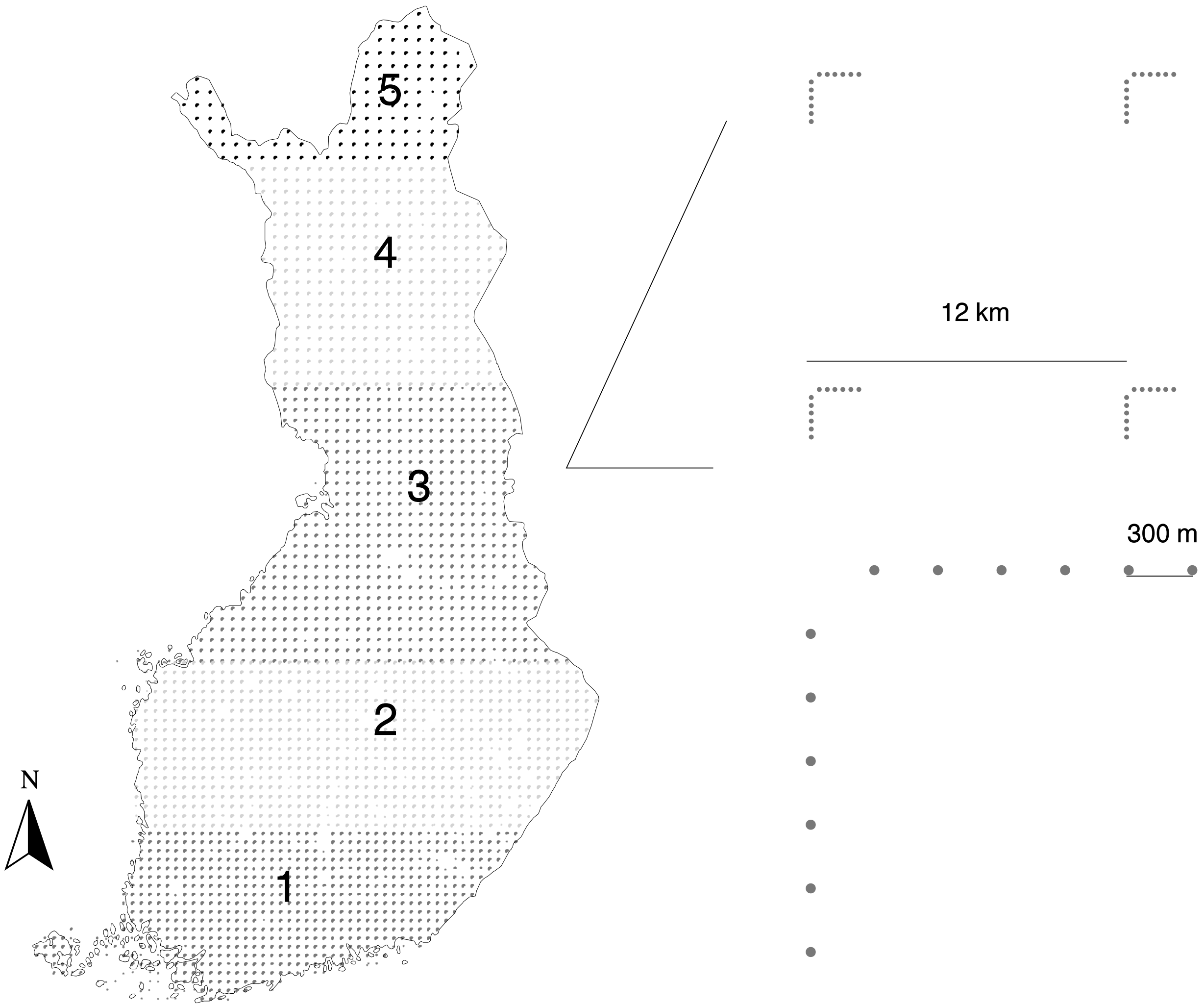
4. A dense NFMA design: NFMA tract distances in latitude and longitude are 1/14 and 1/7 degrees respectively.

5. A dense grid of detached plots with intervals of 4 km $\times$ 4 km, called here a dense Eurogrid.

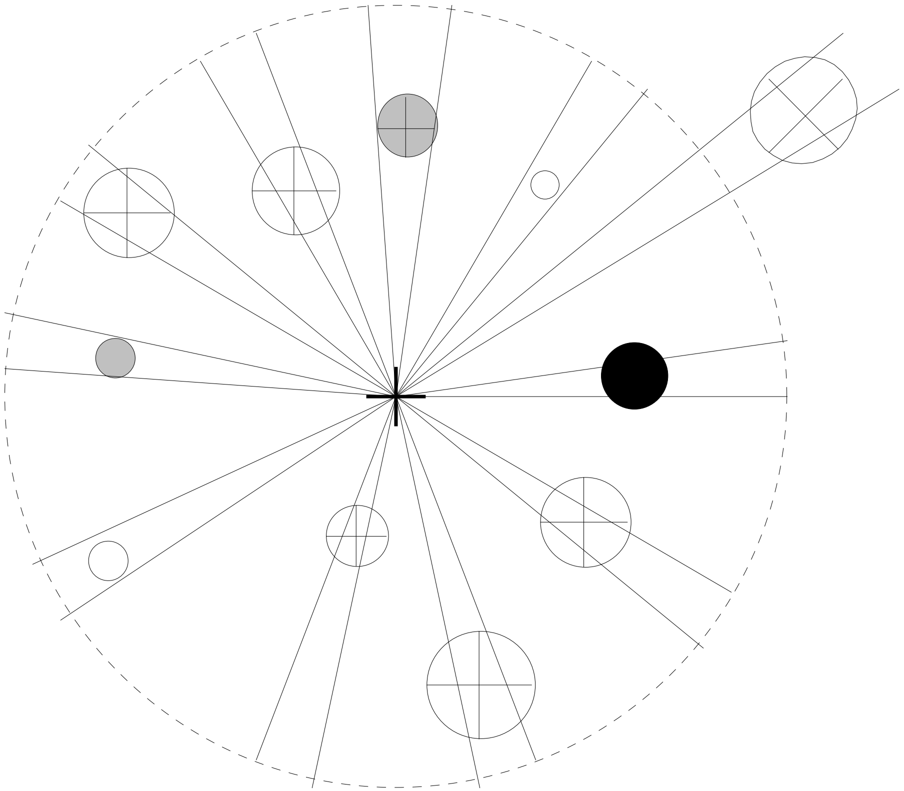
6. A dense cluster design, a cluster consisting of 12 plots located on the sides of a half rectangle with a distance of 300 m apart from each other, and with cluster distances of 10 km $\times$ 10 km (Cluster, no stratification).

7. A dense stratified cluster design the clusters of the plots as in point 3, but the distances between clusters varied in different parts of the country, in the final design, the cluster distances by regions were from South to North 10 km, 10 km, 11 km, 12 km, 13 km and 15 km (Cluster, stratification).

An example of stratified cluster-wise sampling, cluster distances, 10, 11, 12, 13 and km, a cluster in region 3



– Field plot in cluster designs and with detached plots max radius 12.52 m –



Tally tree

Sample tree

A boundary tree - the inclusion checked with measurements - belonging to the sample plot

A boundary tree - the inclusion checked with measurements - not belonging to the sample plot

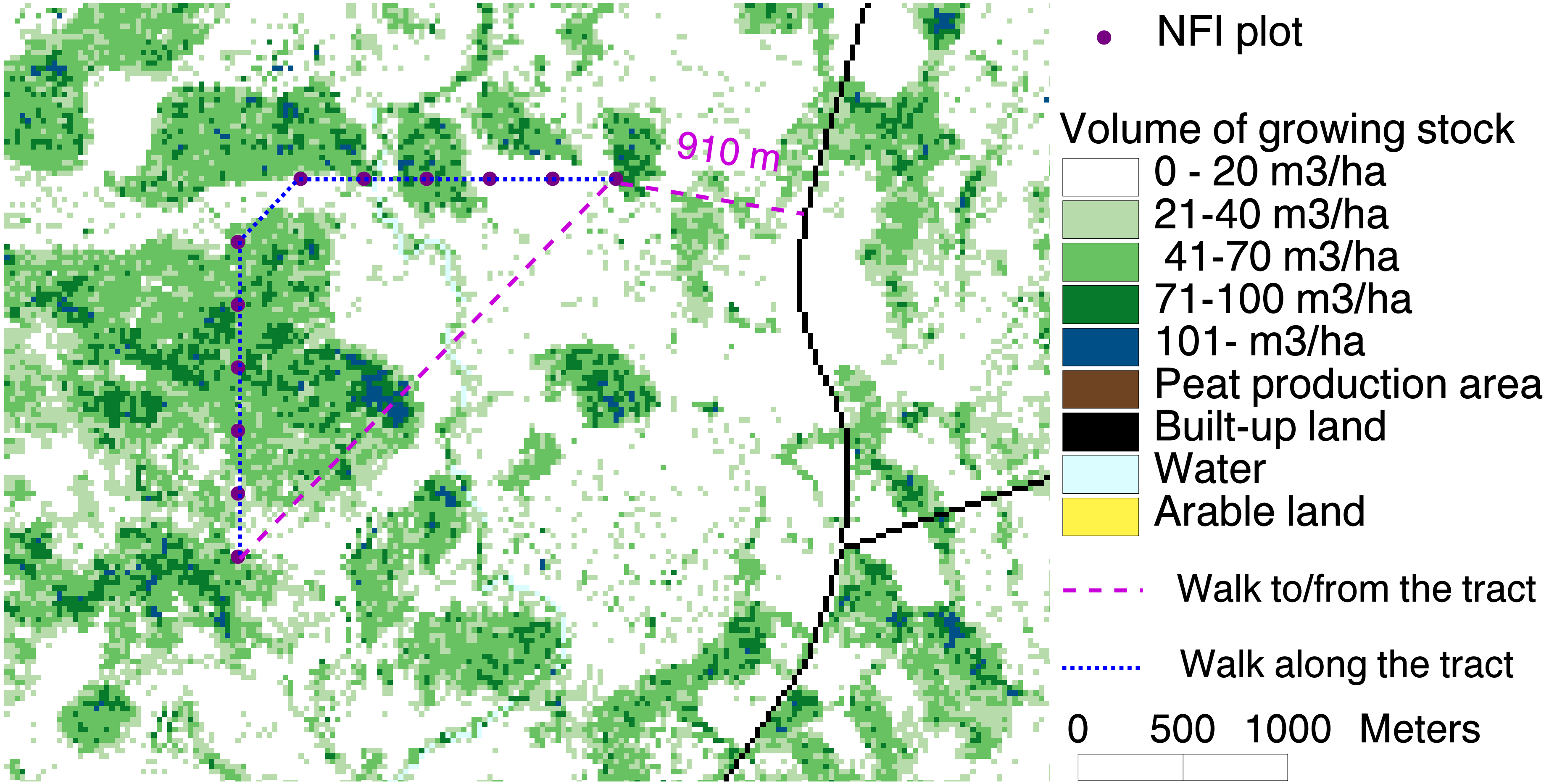
A tree belonging to Bitterlich plot, but outside of the sample plot

– Components considered in determining time –

- Driving to a cluster
- Walking in the field to a cluster
- Measuring of the trees
- Other measurements
- Other actions
- Daily pause

– Cost estimation, cont –

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 NFI9.



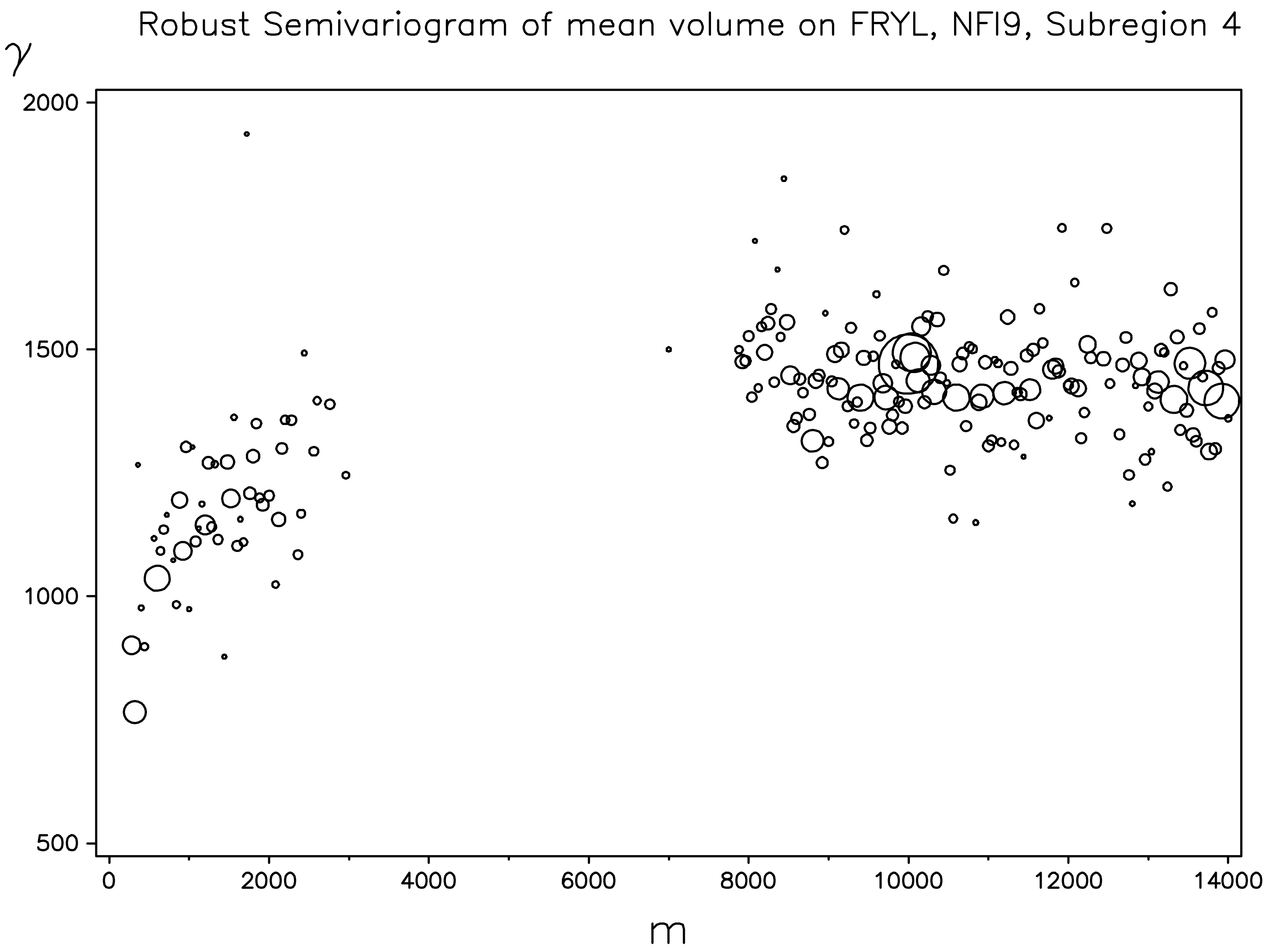
– The error level in the two groups of the designs –

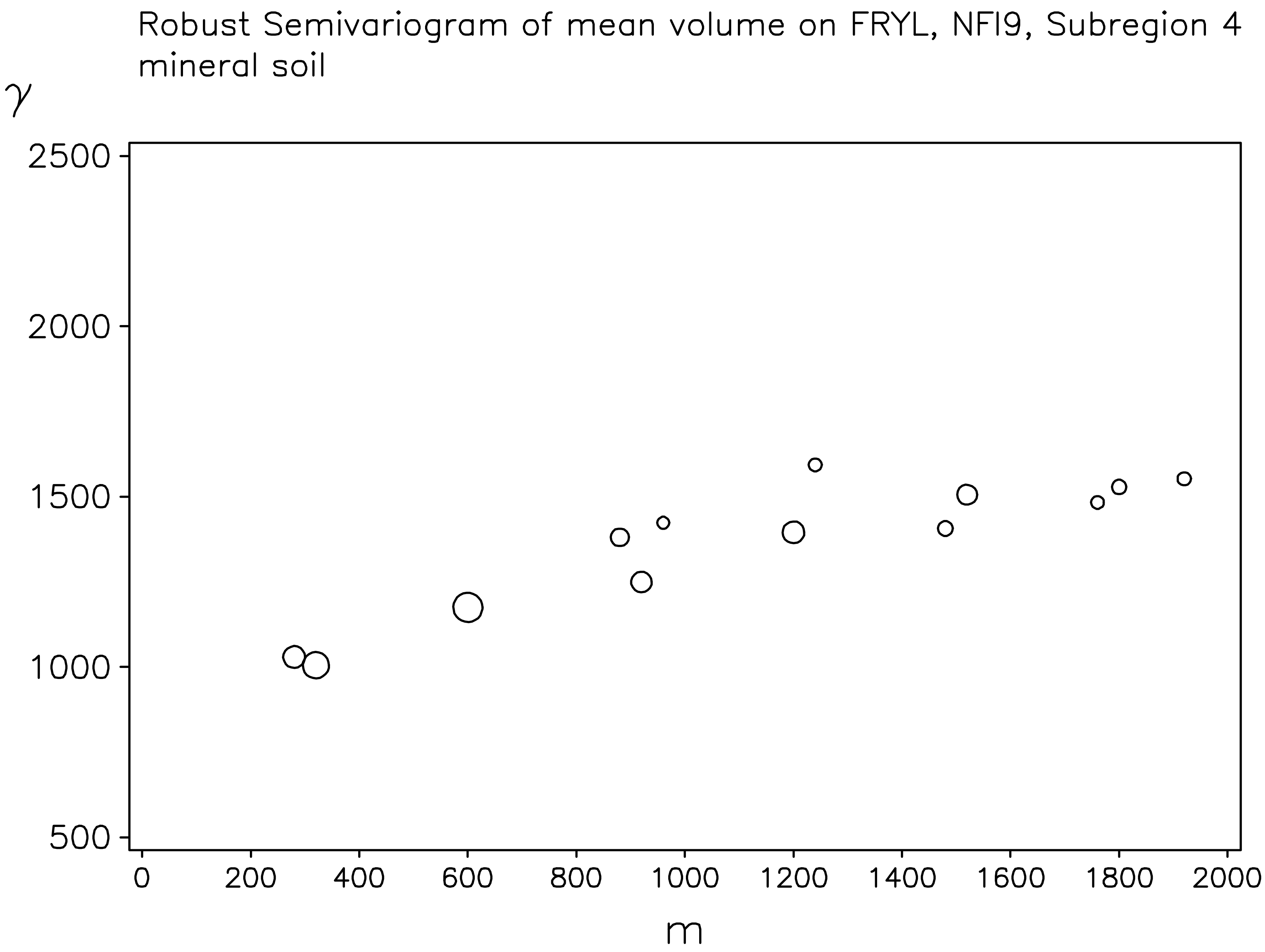
	Volume of growing stock (%) Coefficient of variation (%)		
Designs	ALL	OBL	Saw
Dense	0.60	1.6	1.0
Spare	6.6	21.0	10.1

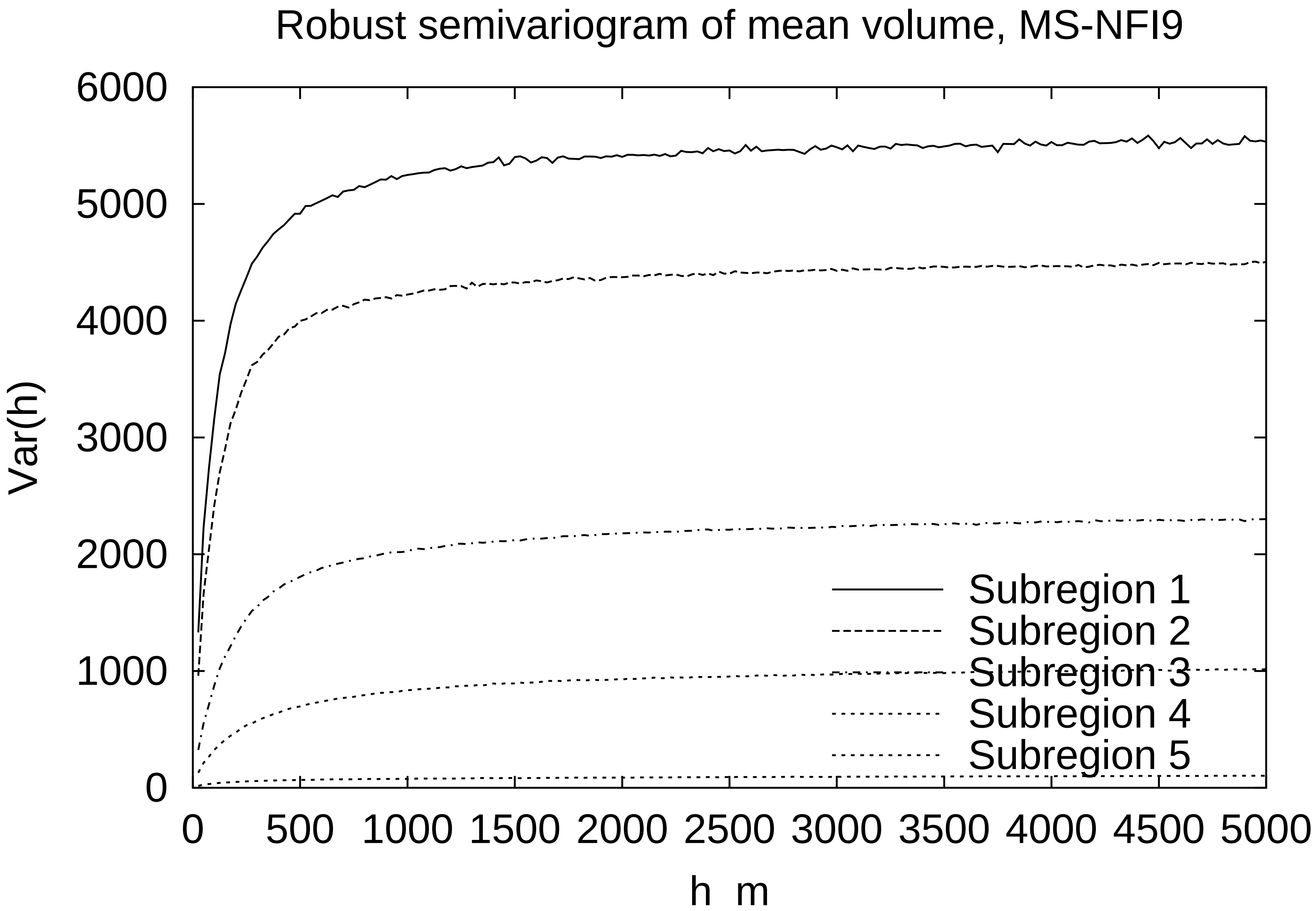
The field crew days and the relative costs to measure the entire country of Finland, covering all land use classes.

Dense design	Crew days	Crew size	Relative time	Relative cost
Stratified cluster design	2773	3	1	1
Non stratified cluster design	3712	3	1.39	1.39
Eurogrid	4502	3	1.68	1.68
NFMA	7712	4	2.89	3.72

Sparse Design	Crew days	Crew size	Relative time	Relative cost
Non stratified cluster design	55	3	1	1
Eurogrid	73	3	1.32	1.32
NFMA	78	4	1.41	1.82


Robust Semivariogram of mean volume on FRYL, NFI9, Subregion 4
γ
2000
1500
1000
500
0
2000
4000
6000
8000
10000
12000
14000
m


Robust Semivariogram of mean volume on FRYL, NFI9, Subregion 4
mineral soil
γ
2500
2000
1500
1000
500
0
200
400
600
800
1000
1200
1400
1600
1800
2000
m


Robust semivariogram of mean volume, MS-NFI9
6000
5000
4000
3000
2000
1000
0
Var(h)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
h m
Subregion 1
Subregion 2
Subregion 3
Subregion 4
Subregion 5

– Variogram and semivariance –

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

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

– Conclusions and possibilities to employ in the Tropics –

- The results are not necessarily directly transferable to the Tropics
  - Variability in the land use structure
  - Within forests variation
  - Further tests are needed
- The methods can be employed in the other ecological zones
- Semivariance studies with data from a similar region, e.g., by country groups or by ecological zones
- The use of possible existing cover map data, e.g., USGS ?
  - Small scale satellite image based preliminary studies are feasible and relatively cheap solutions
- To create standards or blue-print for sampling designs and plot layout, distances between the plots, plot size and shape

– Considerations on latitude / longitude design –

- In theory, each individual in a sampled population should have a positive sampling probability
  - One concern of the NFMA design is that the locations might be widely known

  $\Longrightarrow$ May affect the treatments in the future

- Stratification
  - Already employed in NFMA in some countries
  - Recommendation: adjust sample plot density to the variability of the some core variables

– Considerations on latitude / longitude design, cont –

- Change estimation
- Pros of NFMA
  - Permanent plots, efficient in change detection, helps to meet REDD requirements
- Cons of NFMA
  - A sparse design, small area changes difficult to detect
  - The plot density could be increased and has already been increased in some regions
- Possibilities to increase the efficiency
  - Strips or line transects for land use and land use assessment
  - High or very high resolution satellite images with sampling
    - Satellites exist already
  - Helps also to meet REDD requirements

– Data sources –

- Field data always needed
- Air-borne remote sensing data, air-photos, lidar, as a part of two-phase sampling
- Medium or high resolution space-borne remote sensing data
  - Small area estimation
  - Wall-to-wall maps
  - Output maps for stratification
  - Change detection
- High or very resolution space-borne remote sensing data
  - Relevant with a sampling approach

– Statistical framework –

- Ratio estimates employed in a correct way
- Varying inclusion probabilities
  - The observations further away from the Equator represent a smaller area than those nearer the Equator

  $\Longrightarrow$ equal weighting leads over-weighting of the observations measured further away from the Equator
  - If this is an issue in practice depends on the case country
  - Needs further investigations
  - Easy to take into account in practice, e.g, cluster/plot weights are proportional to the average area of the four closest polygons restricted by the latitudes and longitudes