Increasing the use of remotely sensed data for LULUCF reporting

– should we go design-based or model-based?

Göran Ståhl

Professor of forest inventory

SLU, Sweden

Introduction

- Increased availability of remotely sensed (RS) data offers new possibilities for LULUCF reporting
- Looking at scientifically defendable principles, we can choose to use either of the statistical paradigms ***design-based*** or ***model-based inference*** when making use of the RS data
- This presentation will synthesize preliminary conclusions from development projects that have involved parts of the LULUCF reporting team from Sweden

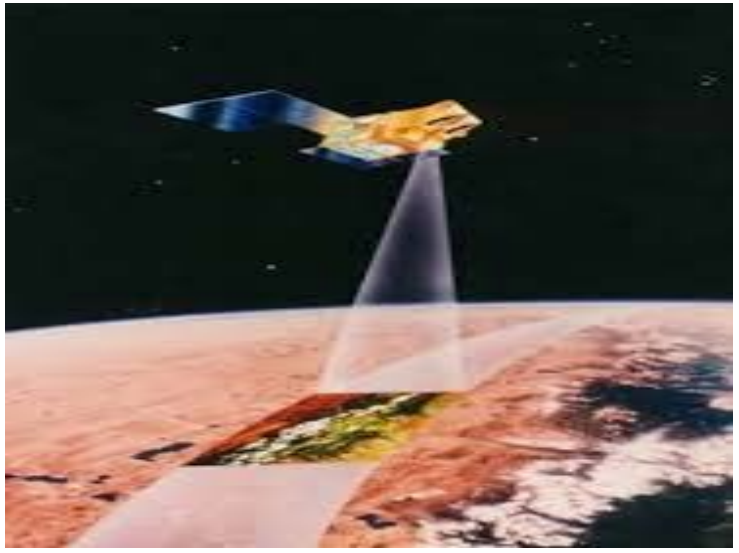
But rather than keeping you on hold, I will tell the answer to the question already now!

The answer is.....

It depends!

So, for the remainder of the presentation I will…

- Briefly introduce the concepts
- Discuss advantages and disadvantages of the two statistical frameworks
- Make some preliminary conclusions regarding their use for LULUCF reporting



Design-based vs. model-based inference

- Statistical inference: How to make scientifically defendable conclusions about entire populations from samples.

  - The populations may be the LULUCF carbon pools within given geographical boundaries
  - We wish to assess how they change and the uncertainties associated with the estimated changes

  - Collect limited amounts of data from the population and build models upon which we rely => model-based inference
  - Collect larger amounts of sample data and apply classical survey sampling methods => design-based inference

Model-based inference with RS data

- Relies on conventional statistical practice, considering values associated with population elements to be random variables
- Uses model relationships between RS data and the variable of interest (e.g. the aboveground biomass on a plot): AGB=f(RS metrics) + e
- Known for a long time but has previously seldom been applied in forest and agricultural monitoring
- ”Rediscovered” in connection with LULUCF reporting and currently gaining increased interest due to limited field sampling requirements

Model-based prediction – the crash course

1) Estimate a model predicting the variable of interest at the level of a plot (pixel), from a sample selected for model fitting
2) Apply the model to every piece of land (pixel) in the population
3) Sum up to obtain a prediction of the total (or mean) for the population
4) Assess overall uncertainty, realising that the major uncertainties are due to:
   - Model specification and application range
   - Model parameter estimation uncertainty
   - Residual error uncertainty

⇒The two latter components can often be quantified and the mean square error for the model-based prediction estimated

Note the difference between formal and informal model-based inference!

- Formal model-based inference utilises a sample from the population to estimate a model, which is subsequently applied
- Uncertainties can be formally assessed
- A plethora of methods are available for *informal* model-based inference, where a model (or a default factor) is applied without strict linkages to the target population
  - E.g., a model or default factor developed from some other region is applied, since no other option is available
  - Uncertainties cannot be formally assessed, unless additional data from the target population are collected

Design-based inference

- Relies on survey sampling practice, considering values associated with population elements to be fixed quantities
- Based on a – typically fairly large – random sample of units from the population
- Applied for a long time in forest and agricultural resource monitoring
- Does not rely on any assumptions, except that observations are made without errors

- RS data and models can be applied as part of design-based inference => model-assisted estimation
- And RS data can also be used for improving sampling designs – important but not further covered in this presentation

Model-assisted estimation with RS data

- Basic theoretical developments in the 1980/1990s, and several large-scale case studies in the 2000/2010s (e.g. Alaska, Norway, …)
- A straightforward way of utilising RS data and models without leaving the paved path of design-based inference
- But requires fairly large samples of field data, in addition to the remote sensing data

Model-assisted estimation – the crash course

1) Apply a model to estimate a population total or mean, as in model-based prediction
   - But in this case we do not trust the model fully!
2) Select a random sample of field plots
3) Compare the model predictions with the field measurements at the level of plots, and formally estimate the total (or mean) *of the differences* between predictions and measurements in the population
4) Model-assisted estimate = Model-based prediction + *estimated difference* (from 3)
5) Uncertainties are assessed according to standard sampling theory

But whatever method we choose there will be errors?

Yes, but we must be aware of them, try to avoid them as far as practicable, and quantify them if they cannot be avoided, as according to the GPG!

Sources of uncertainty (overview)

Method/Error source	Sampling variability	Measurement errors	Model application errors	Model specification errors	Model parameter estimation errors	Model residual errors
Model-assisted estimation	X	X				
Formal model-based prediction		X		X	X	X
Informal model-based prediction		X	X	X	X	X

Some conclusions

- We need a mix of methods to be able to provide comprehensive reports for the LULUCF sector!
- RS data can be used in several ways for improving LULUCF reporting – preferrably through *model-assisted estimation* or *formal model-based prediction*! If possible, *avoid informal model-based methods*!
- Model-assisted estimation has an advantage over model-based prediction, since most of its error sources can be quantified
- Model-based methods encounter problems when reports need to be broken down on categories, such as land-use classes
- Accurate information often goes hand-in-hand with large costs for the data collection. RS data has a potential to improve the accuracy of inventories at given cost levels.
- But everything we are interested in does not correlate strongly with RS data…

Thank you!