Functional Regression Analysis (FRA)

Steen Magnussen[1]
Canadian Forest Service

1. The presentation is a summary from a manuscript currently (June 2016) in review with Remote Sensing of Environment (Elsevier).

Title of submission: **A functional regression model for inventories supported by aerial laser scanner data or photogrammetric point clouds**
Authors: S. Magnussen[a×], E. Næsset[b], G. Kändler[c], P. Adler[c], J.P. Renaud[d], and T. Gobakken[b].

FRA context:

Inventories supported by airborne laser data (ALS) or photogrammetric point clouds (PPC)

Model-assisted and model-based estimators

Note: extensions to hyperspectral data is also possible.

Why FRA?

- ALS and PPC give rise to a potentially very large number of metrics presumed correlated with a study variable (Y) in a population of interest.
- Searching for a model and variables to use in an estimation problem violates the notion of a working model in the design-based paradigm of inference
- FRA affords a formulation of a generic model in advance of sampling, thus satisfying the requirement of a working model and hence model-assisted estimators.

Besides the above rationale one can also argue for FRA on grounds of 'interpretability' and for research into how ALS and AP data connects to our plot data of a study variable.

What is FRA

- *With an explanatory variable X continuous in t:*

$$Y = \int \beta(t) X(t) dt$$

- *With a variable X discontinuous in t:*

$$Y = \sum_{t=1}^{T} \beta_t X_t + e_t$$

Note: the key to success with FRA is to define the yet-to-be-explained ‘t’. It will become clear momentarily.

Estimating the regression coefficients (transfer function)

- Not possible without a set of assumptions and constraints.
- Define X_t as the relative frequencies of canopy heights in the tth bin of canopy height classes.
- Fix number of bins *T*. *T* can be larger than the sample size.
- Assume that $\beta_t > \beta_{t-1} \geq 0$ if $t > t\text{-}1$
- Maximize fit with first-order differences in X. Let η_t , $t = 1,\ldots,T$ denote the coefficients to be estimated.
- Seek a parsimonious model by minimizing the sum of absolute values of η_t (the LASSO).
- Recover the target coefficients $\boldsymbol{\beta}$ from $\boldsymbol{\beta} = \mathbf{A}^{-1}\boldsymbol{\eta}$,
- Where A is a normalized first-order T x T difference matrix
- In model-based inference, use a bootstrap to estimate the covariance matrix of the estimated regression coefficients.

Note: For study variables such as volume per ha, basal area and other attributes related to size and site occupancy of trees the above assumptions appears reasonable.

The estimation 'trick' is to optimize the fit to first-order differences in X(t).
Here we only consider first-order differences, but in other applications we could apply FRA with second-order differences, third-order differences, or a mix of these, all depending on context.

A case study from Baden-Württemberg

Rastatt state forest in the Northern Black Forest with an area of 1012 ha

456 field plots located on a 100 m × 200 m grid and stratified to coniferous (251), deciduous (36), and mixed-wood (169).

Tree attributes such as species, stem diameter (DBH), and height were collected in each sample plot. DBH was measured on a set of four nested (concentric) circular plots.

Volume (aboveground for all parts with a minimum diameter of 7 cm) was calculated by means of volume models developed for the National Forest Inventory, based on taper models (Kublin et al. 2013).

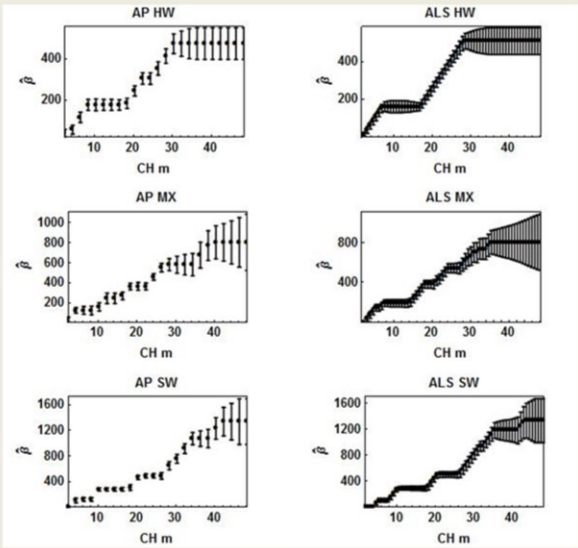
Only trees located inside the forest were measured, and plot areas refer to the area inside the forest.

Note: the submitted manuscript has additional case studies from France and Norway.

Auxiliary variables (X)

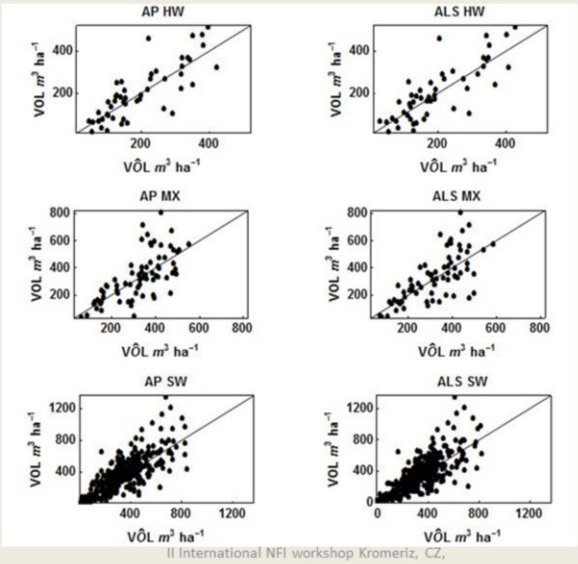
- The auxiliary variables **X** were relative frequency counts of binned canopy heights obtained from aerial photographs **(**AP**)** and ALS first returns.
- PPC data came from the aerial photographs acquired with a Vexcel UltraCamXp camera with a focal length of 100.5 mm. The images overlapped by 80 % along tracks and 60 % across tracks. The ground resolution of the images was 10 cm and the extracted point cloud was interpolated to a regular raster with 1 m^2 resolution.
- ALS data were acquired with a Riegel LMS-Q780 laser scanner at a flying altitude of approximately 600 m above ground. The pulse repetition frequency of 400 kHz and a scan frequency of 150 Hz resulted in a ground point density of approximately 25-50 m^{-2}.
- Frequency counts of ALS and AP canopy heights were binned to T= 24 2.0 m wide classes for AP while the ALS data were binned to T = 63 0.75 m wide classes.

Estimated regression coefficients



Note: similarity across X data sources (ALS and PPC), slight differences across strata (HW = hardwoods, MX = mixed woods, SW = softwoods).

Scatterplot: VOL versus VÔL



Note the apparent divergence of the underlying height-volume relationship in the MX stratum (suggesting a need for models with a change-point). Note also the 'typical' increase in residual variance with increasing predictions. In model-based inference this suggest a need for a variance function for the inference about the variance of a population estimate of the mean. In a design-based inference, the variance of residuals 'integrates' this trend.

Conclusions

- Fit statistics with FRA are comparable to those obtained with selected LiDAR/PPC metrics
- Results are interpretable.
- The model is a true working model and affords model-assisted estimators.
- FRA is less sensitive to plot-size, location errors, and point densities.
- Comparisons across studies are feasible.
- Easy to summarize X for a large area.
- With good DEMs the X from PPC and ALS are equally suitable in forest inventories.

II International NFI workshop Kromeriz, CZ, May 2016

10

Acknowledgements:
The generous supply of data from the institutions of my co-authors is greatly appreciated. Financial support from the Federal Research Institute WSL in Birmensdorf, Switzerland made this presentation possible.

Key references:

James, G.M., Wang, J., & Zhu, J. (2009). Functional linear regression that's interpretable. *The Annals of Statistics, 37*, 2083-2108

Ramsay, J.O., & Silverman, B.W. (1997). *Functional data analysis*. New York: Springer