Predicting soil depth using a survival analysis model

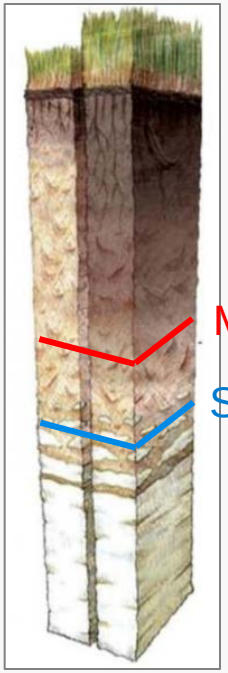
Quentin Styc and Philippe Lagacherie

INRA, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystèmes, Montpellier, France

LISAH

The problem

- Maps of soil depth to bedrock (Lithic or Paralithic contact) are much in demand with the users → included in GSM specs, (*Arrouays et al, 2014*)
- Observations of soil depth are often « right censored »
  - 55% in the French Soil monitoring network (*Lacoste et al, 2016*)
  - 61% in The Languedoc Roussillon Soil database (Vaysse & Lagacherie, 2015)



Maximal observation depth (MOD)

Soil depth to bedrock (SD)

SD > MOD

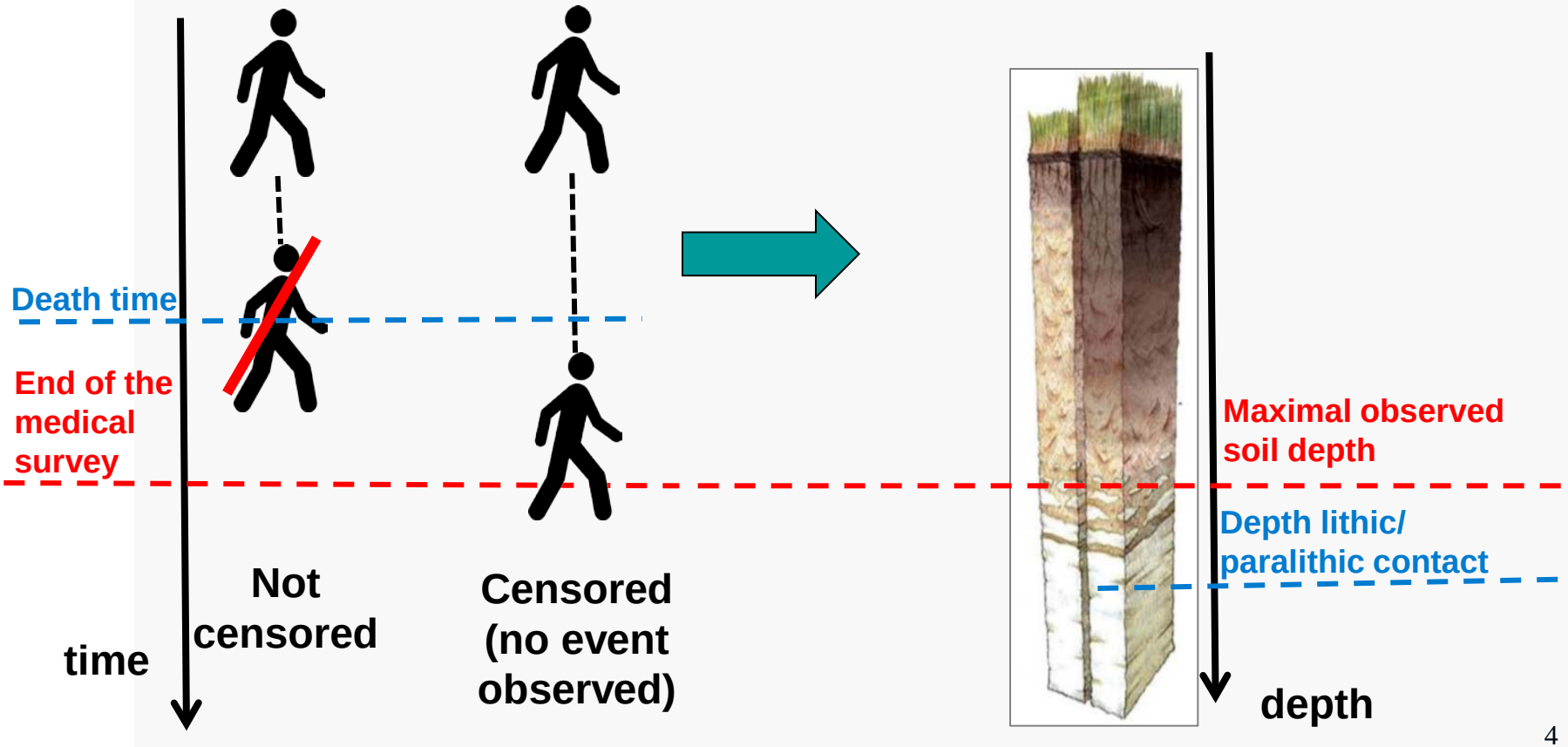
How to use right-censored soil depths in DSM?

Three alternatives

- Deleting right-censored soil depths from the input of DSM models (« just ignore them »)
- Decensoring right-censored soil depths before modelling («soil depth is x cm below »)
- Applying a DSM model tailored for using such inputs

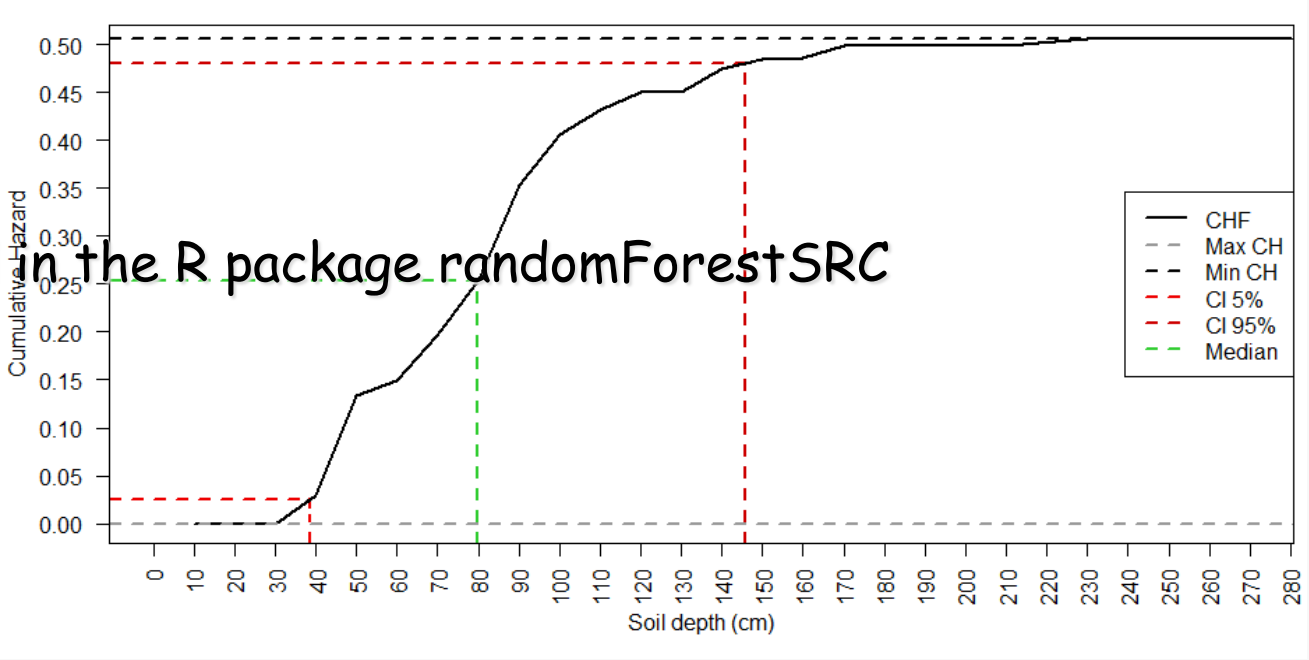
Using Survival Analysis models

- Survival Analysis : branch of statistics for analyzing the expected duration of time until one or more events happen (*Wikipedia*)
- A lot of right censored data in survival analysis

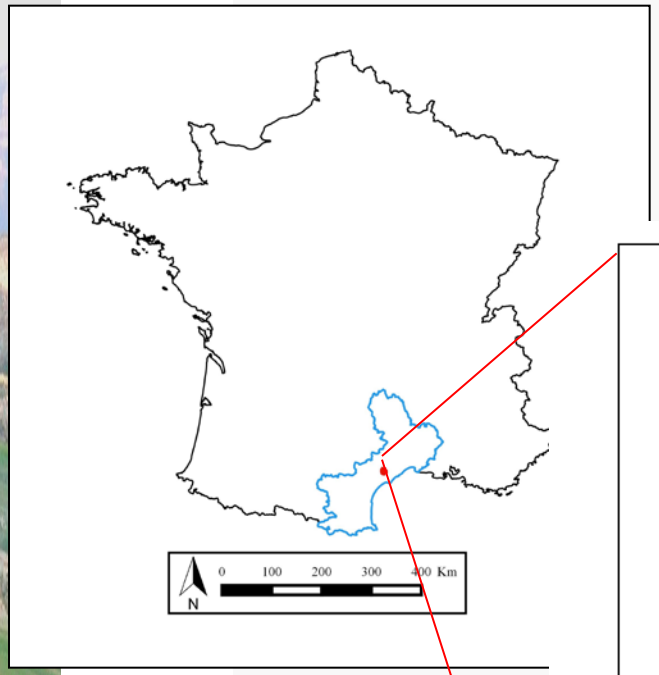


Random Survival Forest *(Ishwaran et al, 2008)*

- Ensemble tree method for the analysis of right censored survival data
- Share the general principle of Random Forests (Breiman et al, 2001)
- Main Specificities
  - The node splitting rule is based on the log-rank test
  - The estimates are survival functions (or Cumulative hazard functions)
- Implemented in the R package randomForestSRC

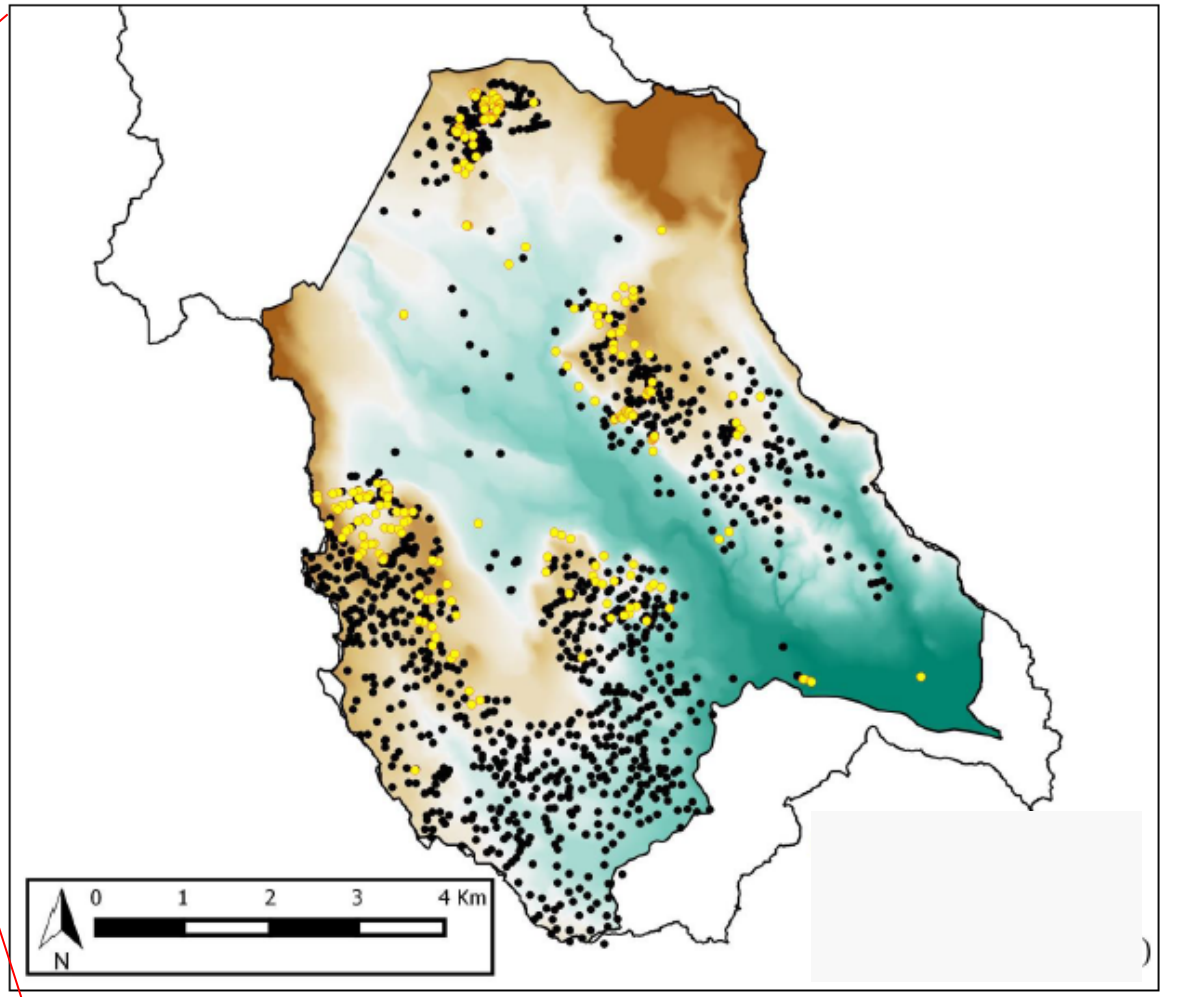


Study area: La Peyne valley (50 km²)



223 sites with exact-valued soil depths

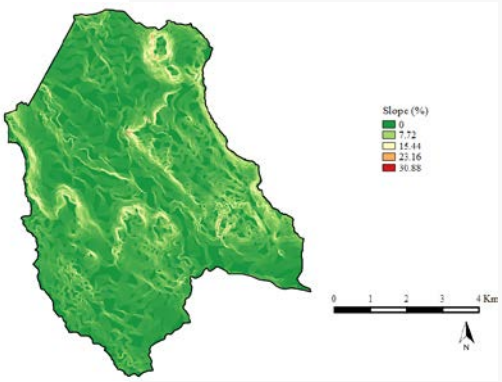
1049 sites with right-censored soil depths



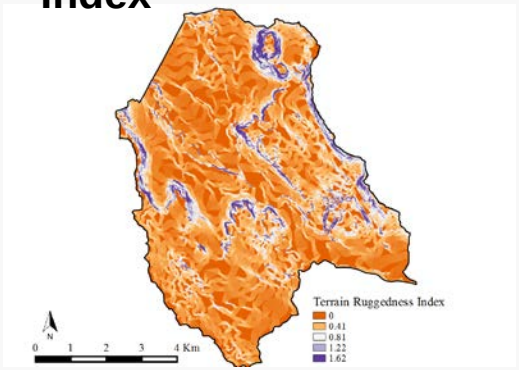
Soil covariates

- 10 meter DEM derived covariates (MrVBF, MrRTF, TWI, TPI, TRI, Slope, plan Curvature, Terrain Roughness)

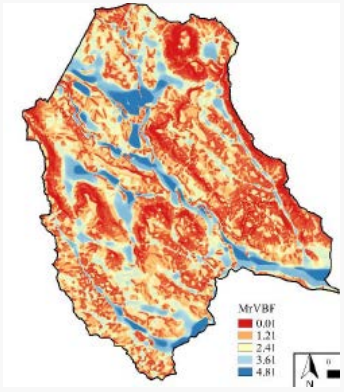
Slope



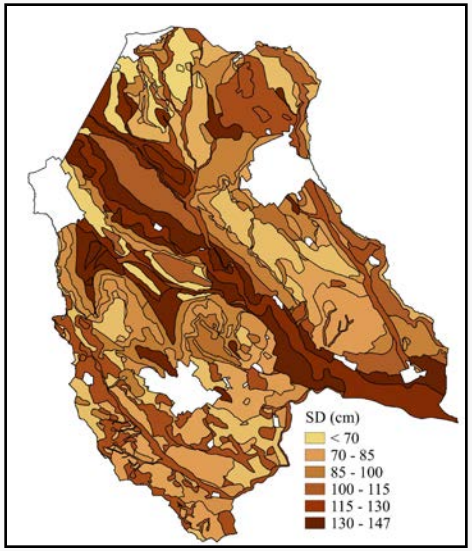
Terrain Ruggened Index



MrVBF



- 1:25,000 soil map



The test

- Comparing Random Survival Forest with Quantile Random Forest (Meinshauzen, 2008) using exact-valued soil depths

- two alternatives using exact valued soil depths only
  - "Just ignore " : calibrate from 223 exact valued soil
  - "Decensore first" (calibrate from 223 exact values + 1049 decensored soil depths)
    - Adding 30 cm to the maximal observation depth (Lacoste et al, 2016)
    - Estimated soil depth = median soil depth over the sites with soil depths exceeding the maximal observation depth

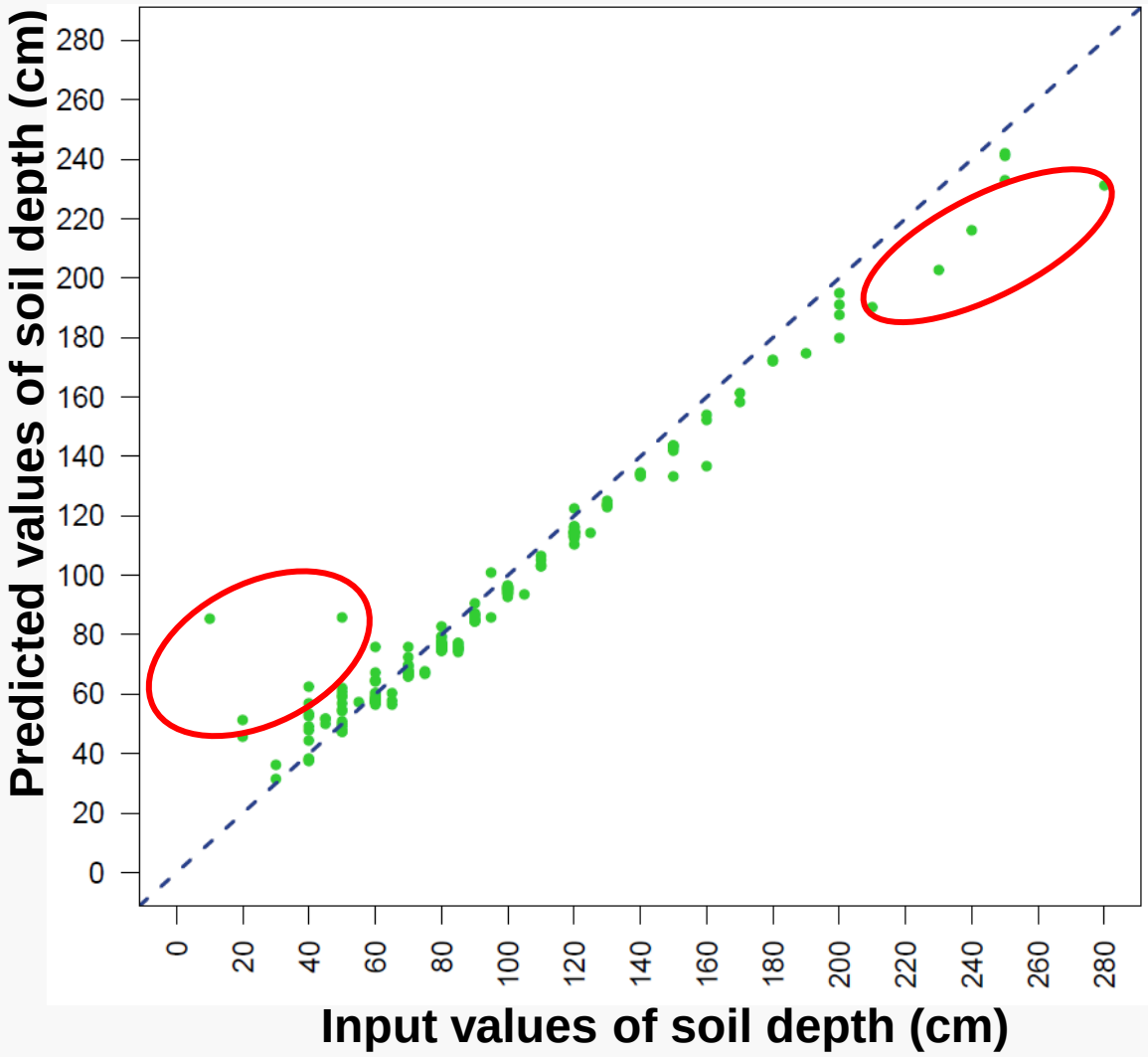
- Validation from 25% of the set of exact valued samples left out (bootstrapped 20 times)

- Visual inspection of the predicted map

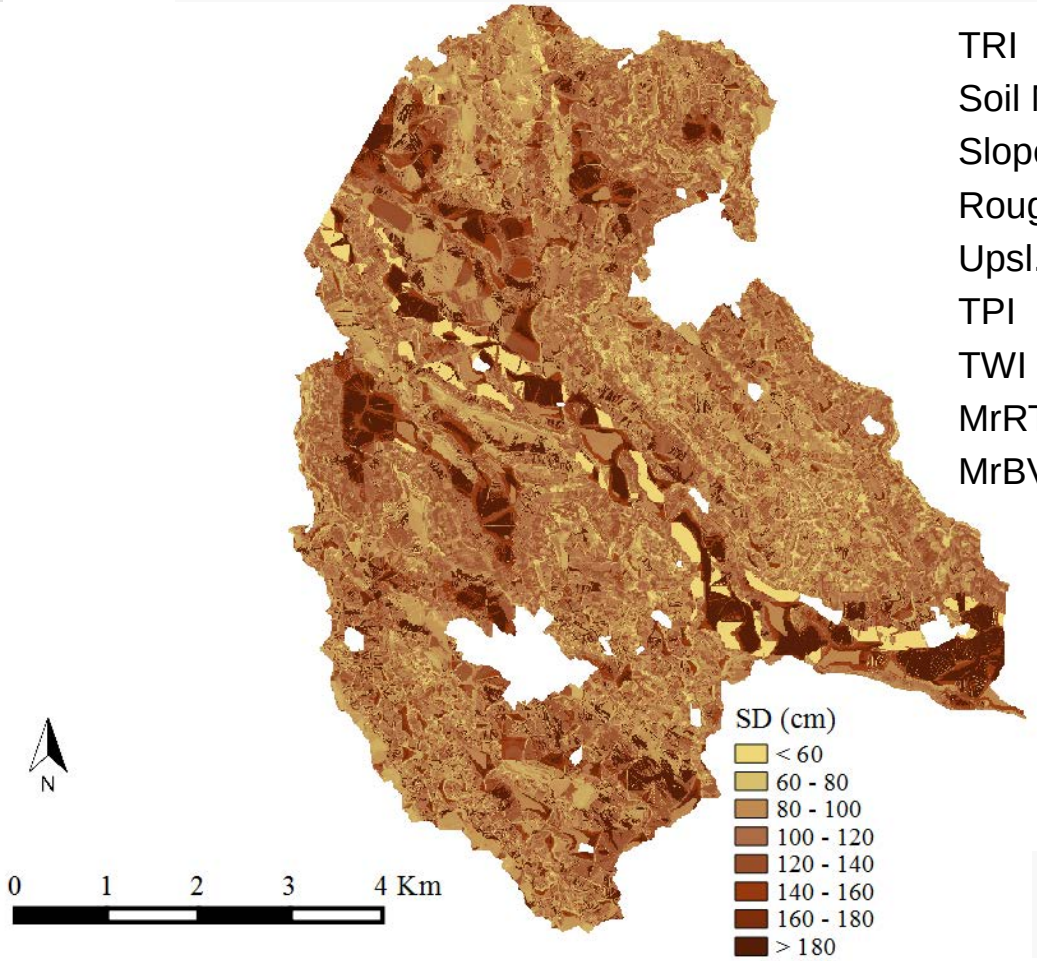
DSM models Performances

SD prediction methods	R^2 *Mean (Std)*	RMSE *Mean (Std)*	Bias *Mean (Std)*	PICP *Mean (Std)*
"just ignore"	0,11 (0,10)	44 (5)	-8 (6)	87 (5)
Decensoring 1	0,15 (0,10)	43 (3)	13 (5)	89 (3)
Decensoring 2	0,15 (0,10)	44 (4)	12 (5)	68 (6)
Rand Surv. For	0,93 (0,04)	12 (3)	-3 (1)	80 (5)

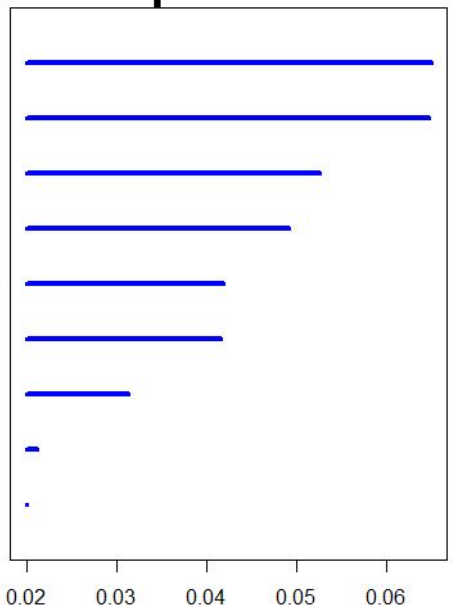
Scatter plot of residuals for RSF (validation)



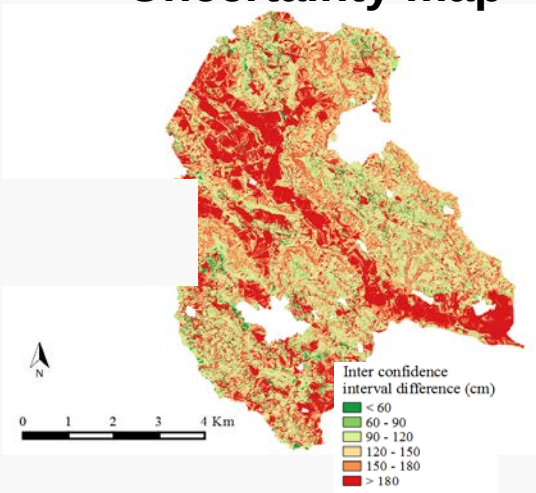
Predicted Map of soil depth obtained with SRF



Covariate importances



Uncertainty map



Conclusions

- Calibrating a DSM model with the sites having exact values of soil depth (1 site/21 ha) obtained weak results in the Peyne valley
- Adding right censored soil depths after decensoring increases the density of sites (1 site/4 ha) but not significantly the prediction performances
- Using survival Random Forest provided good results

→ Need to be tested elsewhere