

# Identifying structure and composition forest types : contribution of classification methods

S. Bonnet<sup>1</sup>, Y. Brostaux<sup>2</sup>, H. Claessens<sup>1</sup> & P. Lejeune<sup>1</sup>

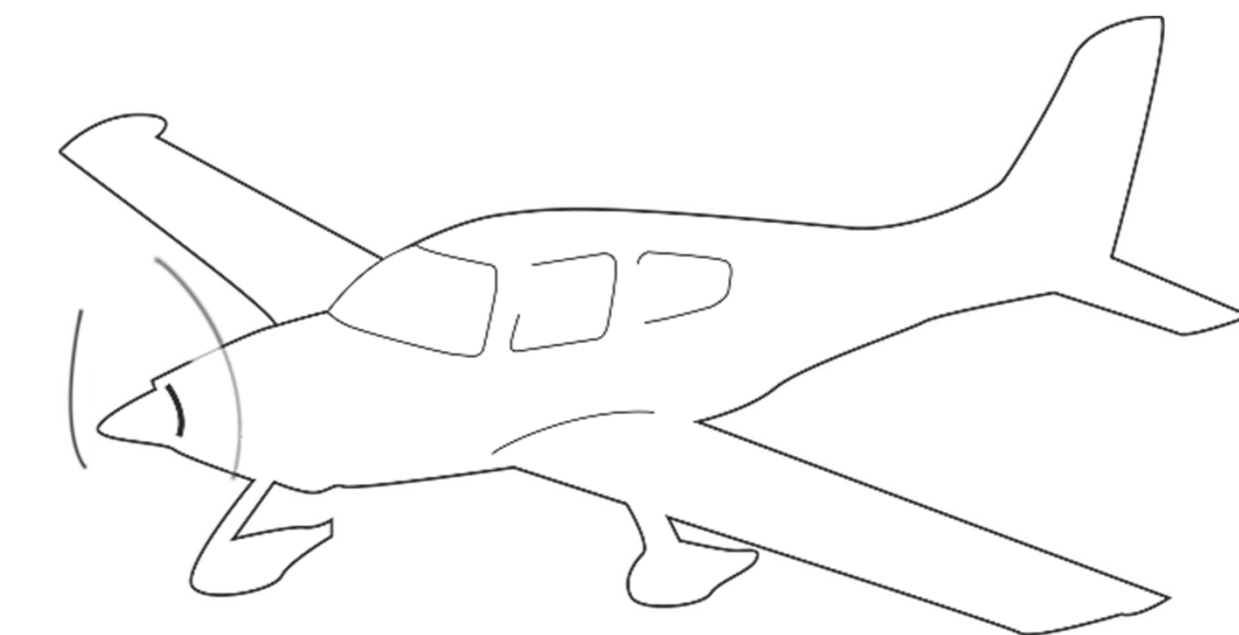
<sup>1</sup> Unit of Forest and Nature Management

<sup>2</sup> Unit of Applied Statistics, Computer Science and Mathematics

s.bonnet@ulg.ac.be

## Context & Objectives

Structure and composition are important for forest and biodiversity management. In this study, we investigate supervised and unsupervised classification methods to classify plots into structure and composition types from LiDAR data. The underlying objectives are: to critically analyze the two typologies used in the study, to highlight relevant LiDAR variables for forest types discrimination and to compare leaf-on and leaf-off data.



## Material & Methods

- Leaf-off and leaf-on LiDAR datasets are used to produce raster metrics at plot-scale. Plots are characterized by LiDAR variables calculated from the rasters (131 variables).
- Reference data are collected by a systematic sampling inventory with visual evaluation of tree girth, in 0,1ha plots. Principle is counting the number of stems by species & by 4 girth classes. Structure types are defined by the percentage of small, medium & large trees (5 types). Composition types are defined by the percentage of oak and beech (4 types).
- Unsupervised methods are used in an exploratory way: analyzing the within-cluster variability and discussing about the 2 typologies relevance. Supervised methods, especially random forests, are first used to identify reliable variables for forest types classification.

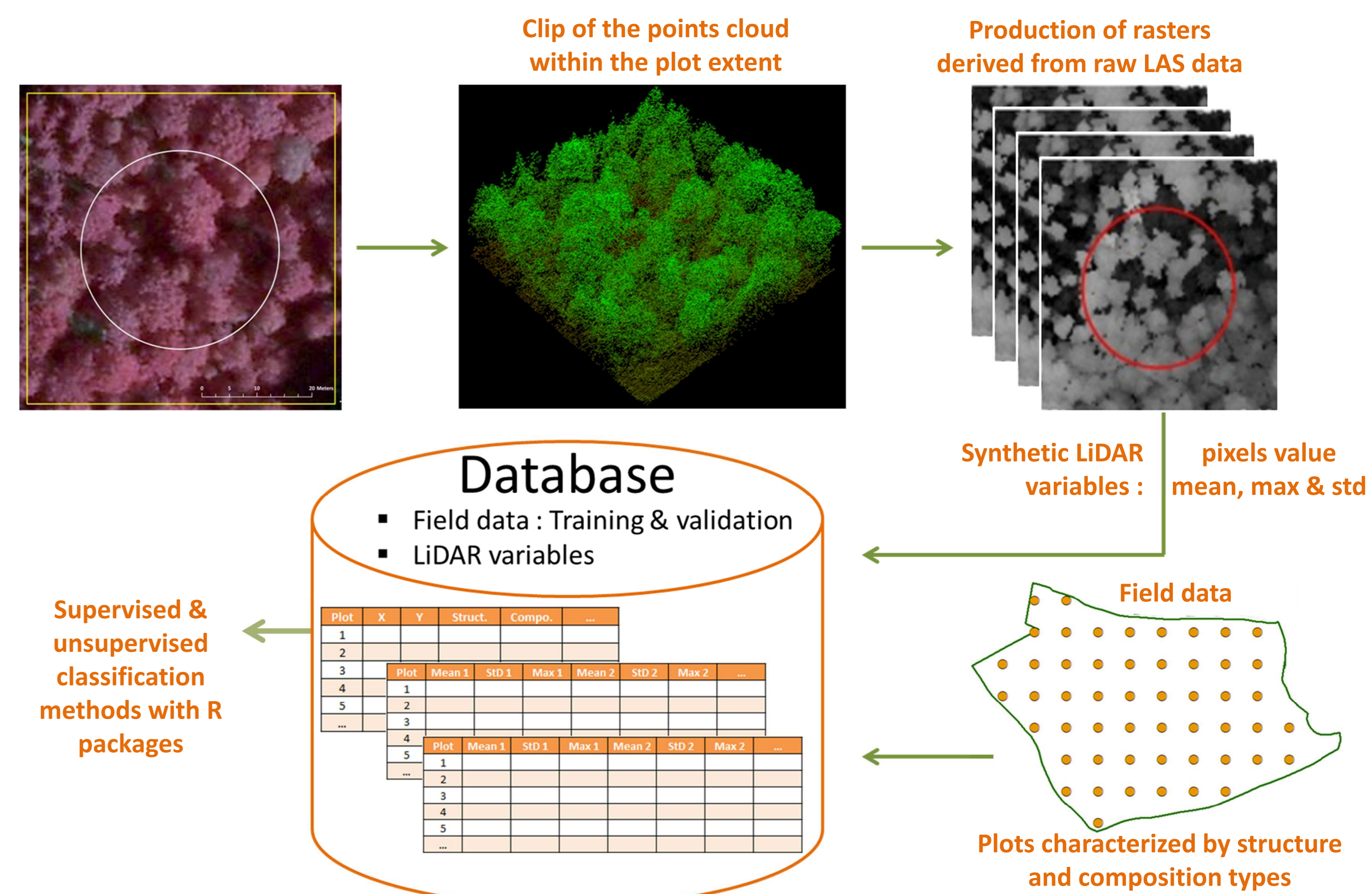
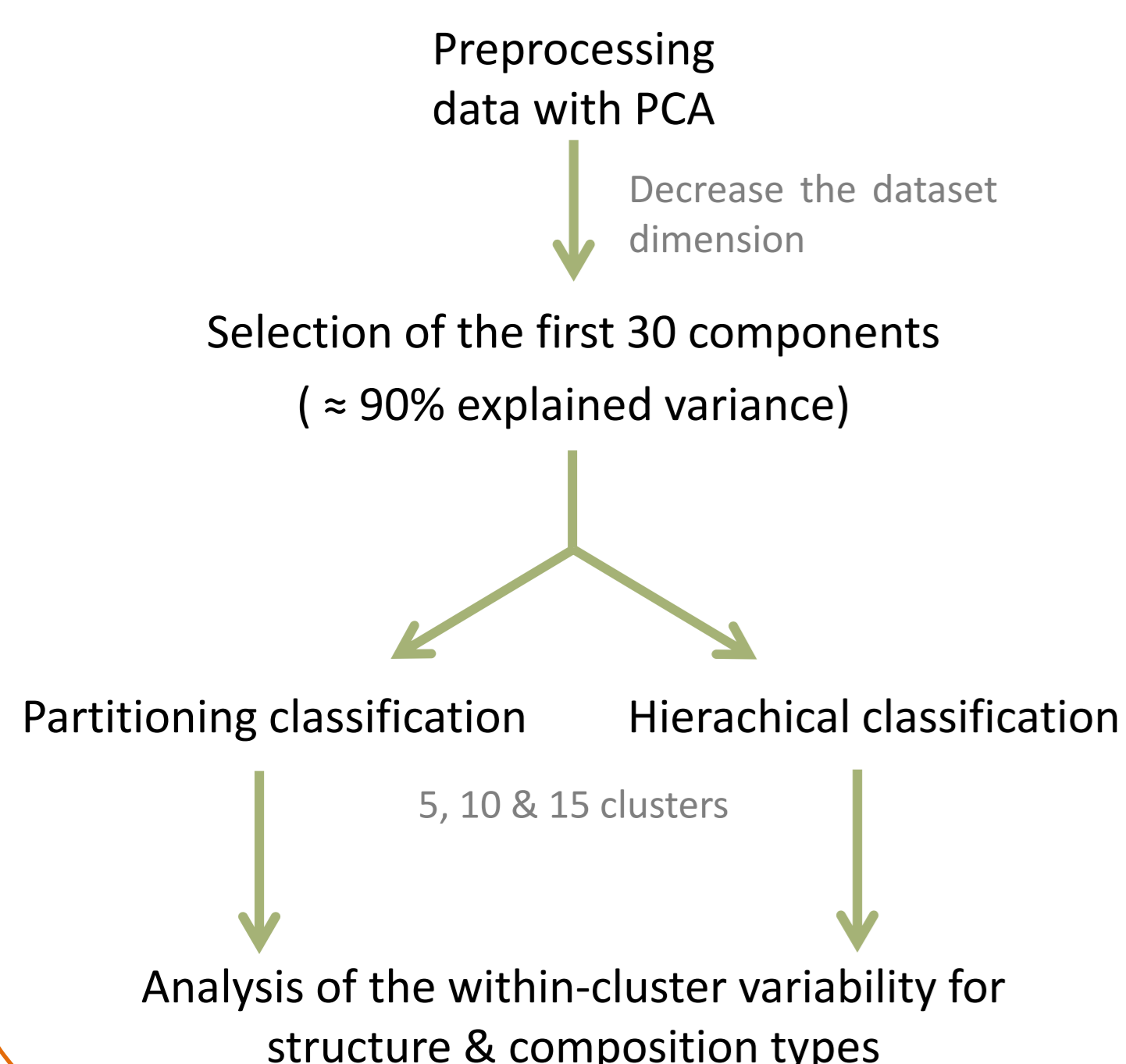


Figure 1 : Implementation of a PostgreSQL database gathering reference data and LiDAR variables. Reference data are collected on 500 plots. For each plot, LiDAR variables are synthesized from values of each raster pixels.

### Unsupervised methods

- Aim** : identification of natural groups within the meaning of LiDAR variables



### Supervised methods

- Aim** : Test the efficiency of supervised data mining methods with LiDAR data to identify forest types

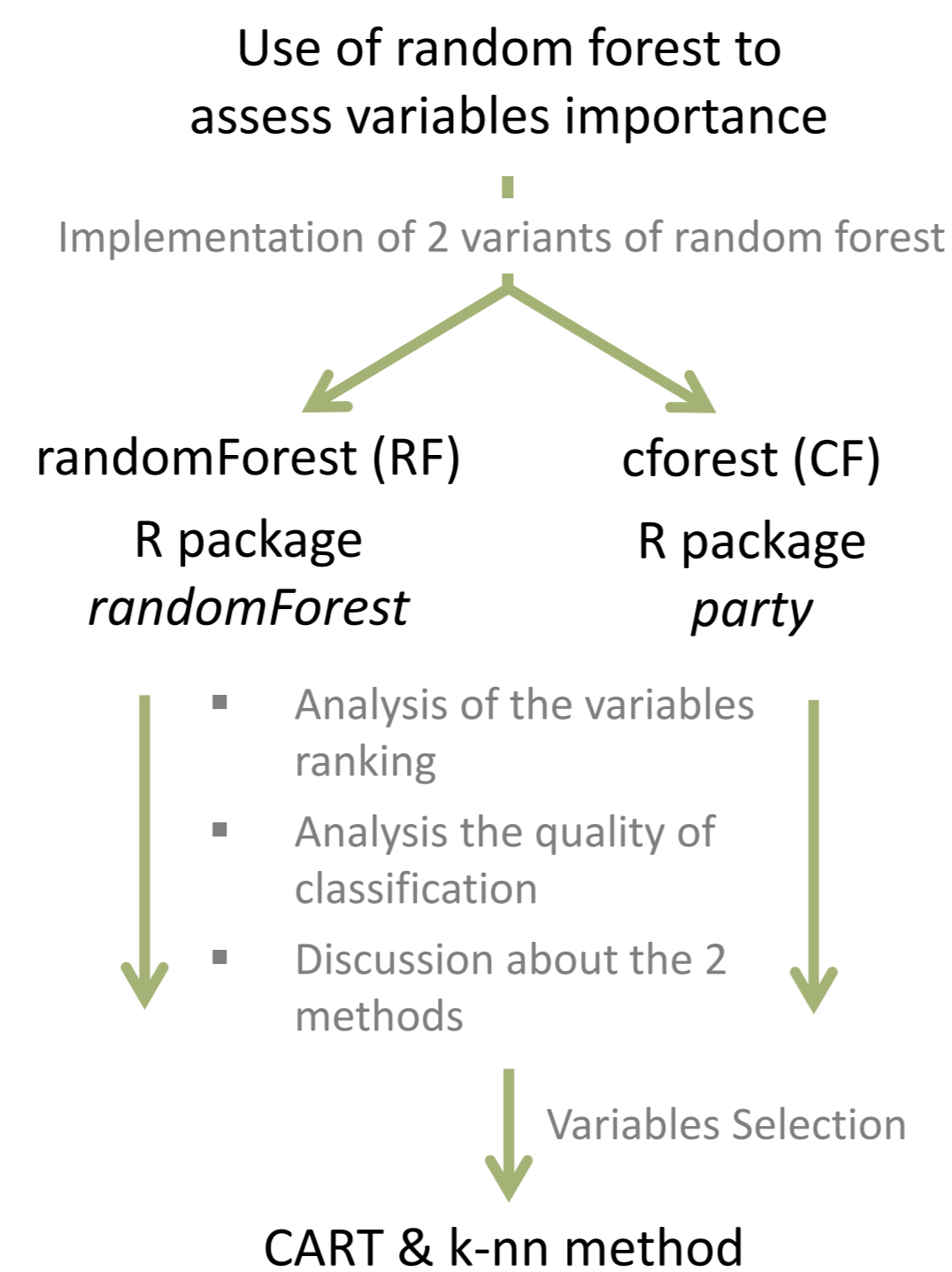


Figure 2 : Schema of the 2 parallel classification approaches to analyze the potential of LiDAR data to identify forest types.

- cforest() is a variant of the classical random forest concept, based on conditional inference for variable selection. cforest was developed to overcome RF typical bias with highly correlated predictors and predictors varying in their scale measurement.
- This typological approach is complemented by an estimation model of basal area (Figure 5).

## Results

- The within-cluster variability shows an important mix of the different structure & composition types (Figure 3).
- RandomForest (RF) results are poor comparing to Cforest (CF) results, especially for structure types (global error ≈ 60 %).
- Global, producer & consumer accuracies are better for composition & for leaf-off dataset (Figure 4).
- For composition, RF seems to better discriminate Beech stands (producer error ≈ 15 % for a global error ≈ 40 %).
- The variables highlighted by RF et CF are different. The 4 most important variables for CF are:
  - For structure** : Max Height of veg. points, 90-95th percentil. of height of veg. points, Std Height of veg. points, % Points between 10 -20 m high.
  - For composition** : Std Height of veg. points, Mean intensity of veg. points, Height coef. var. of veg. points, Std intensity of veg. points.

$$B.A. (m^2/ha) = -0.444 + 2.71*x_1 - 18.87*x_2 + 1.47*x_3 + 17.39*x_4$$

with :  $x_1 = \text{StD}(\text{Height StD of vegetation points})$   
 $x_2 = \text{StD}(\text{Mean height 67th-95th percentile})$   
 $x_3 = \text{Mean}(\text{Mean height 67th-95th percentile})$   
 $x_4 = \text{StD}(\text{Height 90th percentile})$

Figure 5 : Basal area estimation model ( $R^2 = 0.8039$  & Residual standard error = 2.62). Several models were fitted, based raster metrics at plot level and have to be compared and validated.

## Perspectives

- Analysis of the results sensibility & validation with additional plots (also for basal area model).
- Performing k-nn & CART analysis.
- Reflexion on the relevance of the two typologies and possibly development of new ones.
- Injection of new raster variables (selection of the return number).
- Variables calculated directly from the raw point clouds have to be investigated.

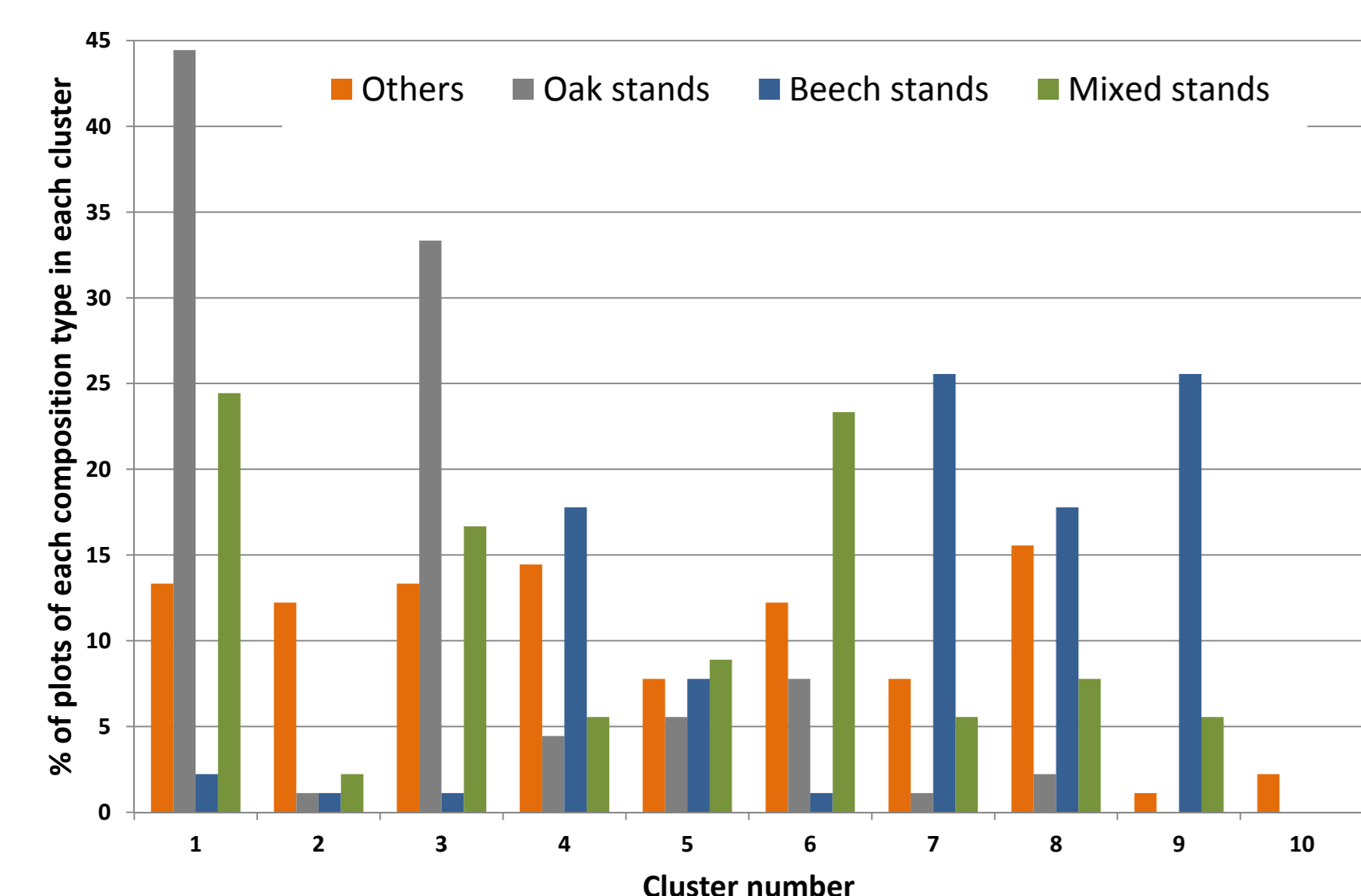


Figure 3 : Within-cluster variability for composition types. The clusters result from a hierarchical unsupervised classification. The dendrogram were cutted to obtain 10 groups.

Leaf-off Composition	Other	Oak stands	Beech stands	Mixed with oak-beech stands
Producer accuracy (%)	77.8	94.4	94.4	62.2
Consumer accuracy (%)	84.3	78.7	85.9	80.0

Leaf-off Structure	Irregular Medium Stems	Irregular Large Stems	Irregular Small Stems	Regular Medium Stems	Regular Large Stems
Producer accuracy (%)	51.4	86.5	86.5	59.5	64.9
Consumer accuracy (%)	76.0	68.1	62.7	75.9	72.7

Figure 4 : Precision for composition (82,2 % global precision) & structure (69,7% global precision) from leaf-off dataset classified with CF. The equivalent RF results give global precisions of 66,11 % & 39,46 %