The principle is, for different LiDAR shoots / assumed as independent, to measure the portions of beam lengths $l_{i,m}$ between successive hits m. Possible censoring leads to introduce binary variables $H_{i,m}$ indicating whether at least one hit occurred or not. Current approaches aim at estimating the dependence of $l_{i,m}$ based on successive hits $m$. In the statistical model, variables $H_{i,m}$ are assumed as independent, to measure the portions of leaf area density (LAD), quantity of interest) to the distribution of beam length $l_{i,m}$ through the statistical model.
where $S_{i,m}$ is a known weight defined as the surface of beam section at shot $i$ divided by the number of echoes. This simplified model does not take into account several effects that may impact either $k$ or beam lengths: leaf orientation, clumping and the nature of hit material: wood vs. leaves, leading to biases or deteriorated uncertainty in estimation. Justifications for the model are to be found in Vincent et al. (2017) and Pimont et al. (2018).

### Principales activités

#### Tasks

**Clumping** is related to spatial aggregation of scatterers at several levels of organization within canopies: crowns, branches, twigs, etc. This is accounted for by space discretization into voxels, followed by local estimations of PAD. However the distributions of observations become then dependent between contiguous voxels, which has to be included into a comprehensive model.

In this project, the aim is to introduce some well-founded statistical method to account for weights and spatial dependencies within a Bayesian framework. A promising approach is offered by hidden Markov random fields. These combine the notions of spatial heterogeneity and dependencies by using latent clusters. The absence of an efficient exact method for estimating posterior distributions and the difficulty to choose the number of clusters led to developing nonparametric Bayesian models with variational approximations, using the VBEM algorithm. These were applied to different problems than LAD estimation (Durand et al., 2021). However, the impact of using weighted data and determining priors by type II likelihood maximization is not well understood in this class of models. The interactions between the weights, choices of priors and concentration of the posterior have to be characterized and assessed using various alternatives. Moreover, the model would have to be refined by including the effect of the distribution of leaf orientation.

**Wood / leaf segmentation** - the discrimination between wood and leaves is required, since both components have contrasted geometric properties (size, volume and orientation of scatterers). Detecting and isolating these components would then allow modellers to apply the previous model to the leaf component only.

To facilitate segmentation, we benefit from some geometric information attached to the position of hits and from radiometric information attached to the intensities of LiDAR returns. Recent results issued from convolutional neural networks applied to dense sets of points from terrestrial laser scanners are quite encouraging (Morel et al., 2020; Wu et al., 2020). However, these results were obtained from low-complexity canopies from denser clouds of points that those acquired from drone laser scanners (DLS); thus the approach will have to be revisited and combined with detection of more specific parts of the trees, e.g., trunk basis.

**Application in ecology:** The different models developed in this PhD project will be applied to repeated DLS measurements acquired in the French Guiana tropical forest, within the PhenOBS program. The latter aims at obtaining maps of the seasonal variations of 3D leaf densities for enhanced characterization of interactions between climate and gas exchange, as measured by flux towers.

**Target profile:** MSc degree in statistics / data science with strong interest in applied modelling. Skills in python / R programming will be appreciated

**Soft skills:** rigor, team work, motivations for problems in ecology

**References:**


### Compétences

**Target profile:** MSc degree in statistics / data science with strong interest in applied modelling. Skills in python / R programming will be appreciated

**Avantages**

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking: two days per week and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included, income tax excluded).