PhD subject proposal

Bayesian models for spatial dependencies and neural networks for leaf / wood discrimination in LiDAR leaf area density estimation.

Context

Covering just 7% of the Earth's land surface, tropical forests play a disproportionate role in the biosphere: they store about 25% of the terrestrial carbon and contribute to over a third of the global terrestrial productivity. They also recycle about a third of the precipitations through evapotranspiration and thus contribute to generate and maintain a humid climate regionally, with positive effects also extending well beyond the tropics. However, the seasonal variability in fluxes between tropical rainforests and atmosphere is still poorly understood. Better understanding the processes underlying flux seasonality in tropical forests is thus critical to improve our predictive ability on global biogeochemical cycles. Leaf area, one key variable controlling water efflux and carbon influx, is poorly characterized. To monitor evolutions of biomass, leaf area density (LAD) or gas exchange, aerial and terrestrial laser scanner (LiDAR) measurements have been frequently used.

LiDAR acquisition from drones located at Paracou – French Guiana (2020).

The 55m tower emerging from canopy is used for gaz exchange quantification (CO2, H2O).

The fast decrease of point density toward the soil is visible.

Grey levels correspond to radiometric intensities of returns.

The principle is, for different LiDAR shoots *i* 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 connecting leaf area density (LAD *k*, quantity of interest) to the distribution of beam lengths $l_{i,m}$ through the statistical model

$$p(l_{1,1},\dots,l_{n,m_n} \mid k) \propto \prod_{i=1}^{n}\prod_{m=1}^{m_n} k^{S_{i,m}H_{i,m}}\, \mathrm{e}^{-kS_{i,m}l_{i,m}}$$

where $S_{i,m}$ is a known weight defined as the surface of beam section at shoot *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)

<u>Tasks</u>

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.

<u>Application in ecology</u>: 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 gaz exchange, as measured by flux towers.

<u>Working environment</u>: You will be based in Inria Grenoble. Short stays in Montpellier AMAP lab (amap.cirad.fr) are foreseen (lab of co-supervisor). This program will be conducted as part of the PhenOBS project funded by the labex CEBA. It brings together a multidisciplinary and international team that combines different approaches (remote sensing, modelling, ecophysiology, citizen science, plant architecture) to document the phenological diversity in tropical forests within a collaborative project.

Supervisors: Dr. G. Vincent (Tropical ecologist and modeller, expertise in LiDAR applied to vegetation), Dr. JB Durand and F. Forbes (Statisticians). Contact: Jean-Baptiste.Durand@inria.fr

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

Soft skills: rigour, team work, motivations for problems in ecology

Remarks: possibility of a stay in tropical forests (accompanying field campaigns)

Durand, J.-B., Forbes, F. Phan, C.D, Truong, L., Nguyen, H., Fatoumata, D., 2021. Bayesian nonparametric spatial prior for traffic crash risk mapping: a case study of Victoria, Australia. Submitted to *Australian & New Zealand Journal of Statistics*. https://hal.inria.fr/hal-03138803

Morel, J., Bac, A., Kanai, T., 2020. Segmentation of unbalanced and in-homogeneous point clouds and its application to 3D scanned trees. *The Visual Computer* 36, 2419–2431.

Pimont, F., Allard, D., Soma, M., Dupuy, J. L., 2018. Estimators and confidence intervals for plant area density at voxel scale with T-LiDAR. *Remote Sensing of Environment* 215, 343-370.

Vincent, G., Antin, C., Laurans, M., Heurtebize, J., Grau, E., Durrieu, S., Dauzat, J., 2017. Mapping plant area index of tropical evergreen forest by ALS. A cross-validation study using LAI2200 optical sensor. *Remote Sensing of Environment* 198, 254-266.

Wu, B., Zheng, G., Chen, Y., 2020. An Improved Convolution Neural Network-Based Model for Classifying Foliage and Woody Components from Terrestrial Laser Scanning Data. *Remote Sensing* 12, 1010.