Convex hull: Another Perspective about Model Predictions and Map Derivatives from Remote Sensing Data

J.P. Renaud[1, 2], A. Sagar[2, 3], P. Barbillon[4], O. Bouriaud[2], C. Deleuze[1], C. Vega[2]
[1]Office National des Forêts, Département RDI,
8 allée de Longchamp, F-54600 Villers lès Nancy, France
Email: {jean-pierre.renaud-02; christine.deleuze}@onf.fr
[2]Institut National de l'Information Géographique et Forestière (IGN), Laboratoire d'Inventaire Forestier (LIF),
14 rue Girardet, 54000 Nancy, France
Email: {cedric.vega; ankit.sagar}@ign.fr
[3] UMR Silva, Université de Lorraine, faculté des Sciences et Technologies
Campus Aiguillettes, 54506 Vandoeuvre Les Nancy
[4]*UMR MIA-Paris, Université Paris-Saclay, AgroParisTech, INRAE* 16 rue Claude Bernard, 75 231 Paris Cedex 05
Email: pierre.barbillon@agroparistech.fr

1. Introduction

In forest inventories as well as in the process of building models, obtaining an efficient sample is a central goal to reach precise estimates of forest attributes (Hawbaker et al. 2009, Frazer et al. 2011, Grafström et al. 2014, Saarela et al. 2015, Bouvier et al. 2019). In a model-based approach, a plots sample must cover adequately the variability of the considered forest attributes in order to minimise prediction error. Different strategies have been proposed to efficiently distribute the field sampling units in the auxiliary space of the remote sensing data (e.g. Hawbaker et al. 2009, Grafström et al. 2014). Some authors have proposed to stratify Airborne Laser Scanning data (ALS) to optimize sampling (Hawbaker et al. 2009, Frazer et al. 2011), and Maltamo et al. (2011) compared different field plot selection strategies in order to optimise models precision.

Interestingly, White et al. (2013) applied convex hull approach to show uncovered forest structures by the field calibration sampling units, since large prediction errors could be associated with model extrapolations, resulting in potentially biased map derivatives. In this research, we use convex hull to identify the proportion of extrapolated pixels, computed their distance to the calibration domain and estimated bias associated to the linear model predictions on an ALS case study.

2. Data and Methods

The study area is based on an ALS flight performed in February 2019 in Northeastern France, which covers a forested area of 18,646 ha, with a pulse emission density of 16 points per m². Within this area, a set of 487 systematic field plots were carried out in the Mouterhouse forest (5,324 ha) during the winter of 2019-2020 (Figure 1). The Mouterhouse forest consisted of broadleaved, mixed and coniferous stands (respectively 43%, 23% and 34% on an area basis) and is representative of the whole ALS area. The main species are Scots pine, sessile oak, beech, Douglas fir and Norway spruce. Diameter and tree species were measured in these calibration plots of 15-meter radius spaced at every ~300 meters. An independent set of 56 plots, located in the same forest area was used for validation.

In order to evaluate the effect of different sampling efforts, the systematic plots grid was thinned by half successively to obtain coarser grids, up to a remaining minimum of 8 plots. The resulting grid sizes obtained were then respectively of 487, 247, 115, 56, 25, 11 and 8 plots.

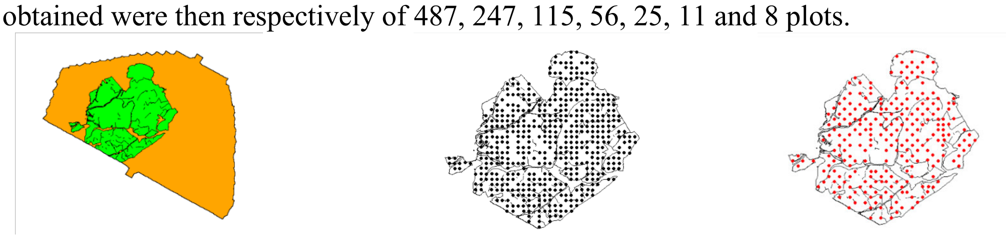

Figure 1: The ALS flight area in orange, with the Mouterhouse forest in green (left) and for illustration, its initial systematic grid (487 plots in black) (middle) and thinned once (247 plots in red) (right).

For 71% of the ALS acquisition area (in orange in Figure 1), no plots were available, giving the opportunity to examine the impact of model application out of its calibration area. Standard area-based metrics were computed from the ALS tiles using the R package lidR (Roussel and Auty 2021). A simple linear model was built to estimate basal area (G), quadratic mean diameter (Dg) and plot density (N). It

Published in: Markus Hollaus, Norbert Pfeifer (Eds.): Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021.
Technische Universität Wien, 2021. DOI: 10.34726/wim.1861
This paper was peer-reviewed. DOI of this paper: 10.34726/wim.1919

included 3 independent variables: mean and standard deviation of all pulse heights and the average slope of the pixels (or of the field plots), computed from the digital terrain model (DTM) (Bouvier et al. 2015).

From these independent variables, convex hulls were computed for all calibration configurations using the geometry package (Habel et al. 2019). It yielded points of the convex hulls and a function allowing to know if a pixel to be predicted belong to the inside or outside space of the hull. The proportion of extrapolated pixels and their distance to the proximal point inside the hull were computed for all pixels of the Mouterhouse forest, as well as for the extended ALS area. Using the validation dataset, it was then possible to compute prediction bias and their distance to the most proximal calibration plot.

3. Results and Discussion

The conventional way to evaluate precision gain associated with sampling efforts is to examine validation RMSEs. In Figure 2, this indicator appears to reach a plateau around 50 calibration plots. With this number of plots however, the proportion of extrapolated pixels in our area of interest (AOI) has not reach a plateau and remains relatively high (ca. 40%). The extrapolation plateau seems rather to occur in the range of 200-300 plots (14-22% extrapolation). Figure 2 also shows a minor difference in the proportion of extrapolated pixels between the two AOI examined. Extrapolation tends to be slightly higher (by ~3%) within the extended AOI, where no calibration plots were present. With a lower amount of calibration plots, the extrapolation distance also tends to increase (Figure 2).

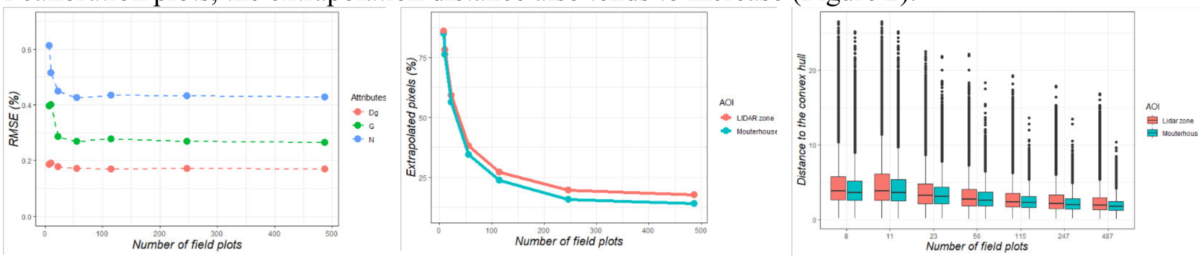

Figure 2: Impact of sampling effort on root means squared errors (RMSE) of validation plots for the different forest attributes examined (left) and on the proportion of pixels located outside the convex hull according to the area of interest (whole ALS or only Moutherhouse) (middle). Boxplots of the distances to the hull are also presented (right).

As large prediction errors might be associated with model extrapolation, Figure 3 presents for each forest attributes examined, the bias obtained with the validation datasets as a function of their distance to to the most proximal calibration plot. Large positive and negative bias were observed (from -303% to +161%). However, a negative trend appears to occur as a function of distance (a distant position being associated with an overestimation of the model). This trend appears to be low for Dg, but high for G and N. This result indicates that it is not only possible to identify pixels that are extrapolated, but also suggest the possibility to correct this result based on the relationship between bias and extrapolation distance.

In this study, we used a simple model, but a preliminary work also showed that more complex models are generating larger proportion of extrapolated pixels. This is certainly associated to the curse of dimensionality. Therefore, the approach used in this study could be considered as an interesting tool to compare competing ALS models.

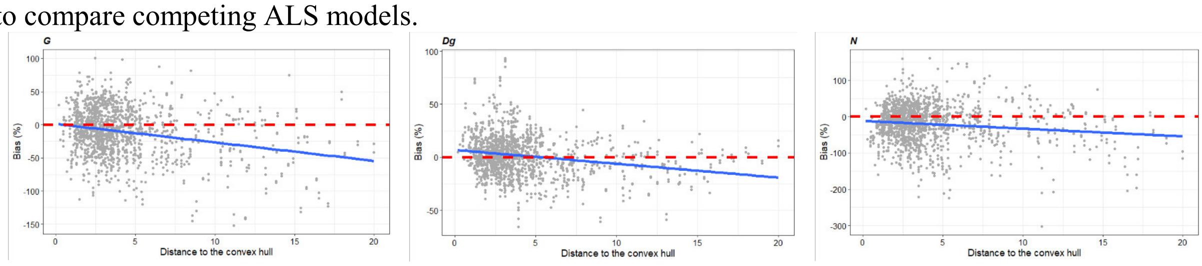

Figure 3: Relative bias for the different forest attributes examined associated to the extrapolated validation plots as function of their distance to to the most proximal calibration plot. (Attributes are basal area (left), quadratic mean diameter (middle), and plot density (right.)

Extrapolated pixels are easily identified using this approach and can then be mapped. Therefore, the spatial distribution of these extrapolated pixels (mapped) could provide additional information that require further attention. This information could reveal important aspects for forest managers, indicating areas where model predictions are out of the calibration domain.

Finally, as Knn is frequently used with ALS, the use of convex hull could also be considered as an interesting tool for further investigations, since this modelling method is also known to be sensitive to extrapolation.

4. Conclusions

The study showed the possibility to identify (and afterward map) pixels that are extrapolated using a convex hull approach. It also showed that prediction bias tends to increase with distance to the calibration domain and that low amount of calibration plots tend to increase these distances. Convex hull is certainly an interesting tool to evaluate model "representativeness" over an AOI and could then serve to compare models in a phase of model selection. It could also serve at the stage of sampling effort determination. Further studies are nevertheless required to evaluate this tool to correct extrapolation bias.

Acknowledgements

We acknowledge the Laboratoire Interdisciplinaire des Environnements Continentaux (*CNRS et l'Université de Lorraine*) for supplying the ALS data and the Agence Territoriale de Mulhouse (*ONF*) for the field data. We thank Steen Magnussen for enlightening discussions, as well as Brad Barber and David Sterratt for their help with the geometry package. We thank Alain Munoz, Flavien Lamiche, and Vincet Pérez (ONF) for their help in preparing and supplyina data. ONF Département RDI is supported by the French National Research Agency (*ANR*) as part of the "Investissements d'Avenir" program (ANR-11-LABX-0002-01, Lab of Excellence ARBRE). Ankit Sagar received the financial support of the French PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE, through the project Impact DeepSurf.

References

Bouvier M, Durrieu S, Fournier RA, Renaud JP. 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. Remote Sens Environ 156: 322–334.

Bouvier M, Durrieu S, Fournier RA, Saint-Geours N, Guyon D, Grau E and Boissieu F. 2019. Influence of Sampling Design Parameters on Biomass Predictions Derived from Airborne LiDAR Data. Can J Remote Sens 45: 650–672. https://doi.org/10.1080/07038992.2019.1669013

Frazer GW, Magnussen S, Wulder MA, Niemann KO. 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sens Environ 115: 636–649. https://doi.org/10.1016/j.rse.2010.10.008

Grafström A, Saarela S, Ene LT. 2014. Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Can J For Res 44: 1156–1164. https://doi.org/10.1139/cjfr-2014-0202

Habel K, Grasman, R, Gramacy RB, Mozharovskyi P and Sterratt DC. 2019. geometry: Mesh Generation and Surface Tessellation. R package version 0.4.5. https://CRAN.R-project.org/package=geometry.

Hawbaker TJ, Keuler NS, Lesak AA, Gobakken T, Contrucci K, and Radeloff VC. 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. J Geophys Res G Biogeosciences 114: 11 p. https://doi.org/10.1029/2008JG000870

Maltamo, M, Bollandsas, OM, Næsset, E, Gobakken, T, and Packalen, P. 2011. Different plot selection strategies for field training data in ALS-assisted forest inventory. Forestry, Vol. 84 (1): 23–31. doi:10.1093/forestry/cpq039

Roussel JR and Auty D. 2021. Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 3.1.1. https://cran.r-project.org/package=lidR

Saarela, S, Schnell, S, Grafstrom, A, Tuominen, S, Hyyppa, J, Nordkvist, K, Kangas, A, and Stahl, G. 2015. Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. Can. J. For. Res. 45(11): 1524–1534. doi:10.1139/cjfr-2015-0077

White JC, Wulder MA, Varhola A, Vastaranta M, Coops NC, Cook BD, Pitt D, Woods M. 2013. A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. Canadian Forest Service, 50 p.