# QC-SF: Improving Computer Vision for Airborne LiDAR Point Clouds of Boreal Forests with Quebec Simulated Forest Dataset

Olivier Stocker, Reza Mahmoudi, Omid Reisi Gahrouei, Thierry Badard and Eric Guilbert
CRDIG, Université Laval, Québec, Canada

WACV 2026 TUCSON, AZ 3/6 - 3/10

## QC-SF : Québec-Simulated Forest

Québec-Simulated Forest (QC-SF) is a large-scale, synthetic airborne LiDAR point cloud dataset of Québec Boreal forest. QC-SF was created to solve the critical lack of labelled airborne data for boreal forest.

### Large-scale

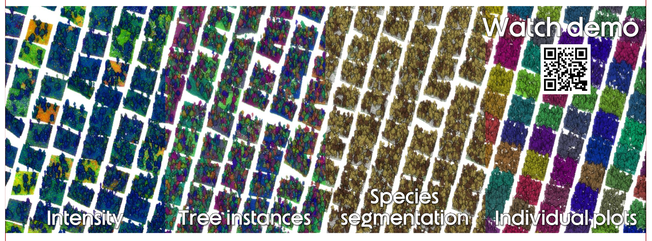
QC-SF provides around 1.7 billions fully annotated points of a virtual boreal forest of Québec region. They were scanned with a virtual airborne LiDAR flying over 60,000 plots of 400m², equivalent of 24km².

### Informative

Plots are highly diverse, both in structure and in content. The three main parameters are tree density, species mingling and low vegetation content. We provide classical scan information as well as annotations for 3 vision tasks.

### Realistic

Every component used to generate the virtual forest is real or backed-up by previous literature. QC-SF is designed with the minimal sim-to-real gap possible.



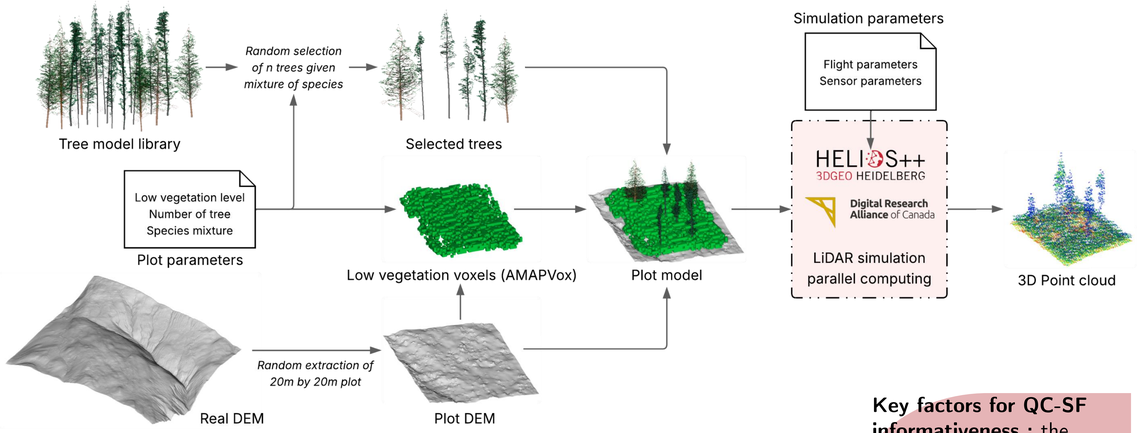

### QC-SF specifications

- Point cloud density : 50pts/m2
- XYZ coordinates, intensity, number of returns and return number
- Labels for 3 visions tasks
  1. Species semantic segmentation
  2. Individual tree segmentation
  3. Wood/Leaf semantic segmentation

QC-SF Dataset open source : **https://qc-sf.org**

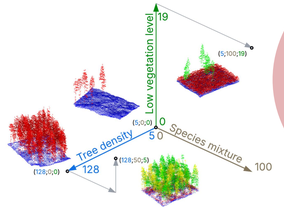
## Simulation Methodology

This simulation pipeline has been used separately for both train and test sets. Both sets have separate ensemble of models and parameter spaces to avoid data leakage. Platform noise was not integrated into the simulation, to avoid baking a single noise representation.




### Key factors for QC-SF realism :

- Ground : 2km² of real LIDAR DEM tiles from LidarBC
- Trees : model library from Côté et al. [1] composed of two species : Black Spruce and Balsam Fir.
- Low vegetation : Procedural generation based of measurements by Liu et al. [2]
- Intensity : Reflectance values for material properties from our field measurement and [3,4,5]



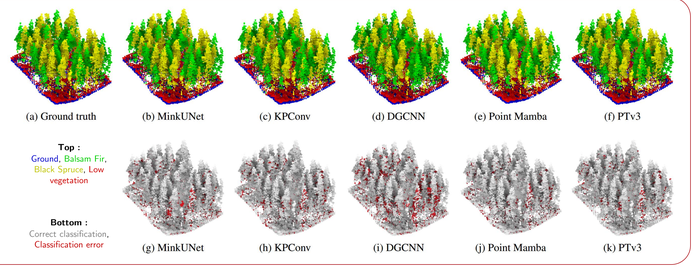

### Key factors for QC-SF informativeness :

the parameter vector space. To ensure plot diversity, we designed a 3-dimensional plot parameter space. Each generated plot occupies a unique position in this space, defined by its specific combination of tree density, species mixture, and understory vegetation level.

## Benchmark

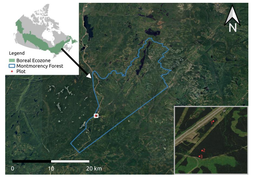
All models were trained from scratch with the QC-SF train set, using an equivalent data augmentation pipeline. Following table and figures are inferences on QC-SF test set.

| Method | OA | mIoU | $IoU_{Ground}$ | $IoU_{AB}$ | $IoU_{PM}$ | $IoU_{Low_veg}$ |
|---|---|---|---|---|---|---|
| MinkUnet | 93.47% | 87.26% | 84.12% | 89.40% | 90.80% | 84.70% |
| KPConv | 93.09% | 86.51% | 83.79% | 89.65% | 89.39% | 83.21% |
| DGCNN | 91.15% | 83.20% | 80.96% | 87.04% | 85.12% | 79.67% |
| Point Mamba | 94.48% | 89.23% | 85.74% | **91.47%** | **91.39%** | 88.35% |
| Point Transformer v3 | **95.61%** | **91.66%** | **93.23%** | **91.40%** | 90.81% | **91.20%** |

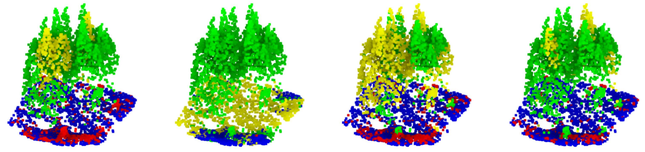



## Sim-to-real transfer

To create a reference dataset, we conducted a field inventory and a high-density ULS survey of three 100m2 plots in Québec boreal forest of Montmorency. After down-sampling to ALS density, real dataset is composed of 102 clouds for train set, and 42 for test set. With this data we trained three PTv3 models from scratch using QC-SF (Sim-only), the real only (Real-only) and a mix of QC-SF and real (Mix). Results below are inferences of the respective networks on the mixed-species plot clouds we kept as test.




| Metric | Sim-only | Real-only | Mix (Sim+Real) |
|---|---|---|---|
| OA | 59.59% | 62.57% | **78.98%** |
| mIoU | 26.33% | 46.39% | **52.12%** |
| $IoU_{Ground}$ | 26.64% | 79.60% | **81.23%** |
| $IoU_{AB}$ | 77.77% | 44.10% | **75.19%** |
| $IoU_{PM}$ | 0.55% | 19.79% | **20.62%** |
| $IoU_{Low_V}$ | 0.36% | **42.05%** | 31.44% |



(a) Ground truth (b) Sim-only (c) Real-only (d) Mix (Sim+Real)

Legend : Ground, Balsam Fir, Black Spruce, Low vegetation

## Conclusion

1. **QC-SF Dataset:** Introduced a large-scale synthetic airborne LiDAR dataset featuring 60,000 fully annotated plots representing the Eastern Canadian boreal forest.
2. **Architecture Benchmark:** Identified Point Transformer v3 as the leading 3D deep learning architecture for forestry semantic segmentation.
3. **Sim-to-Real Success:** Proved that augmenting limited real-world scans with QC-SF synthetic data increases performance by ~6% mIoU.

## Contact and References

Code open source on : **https://github.com/NanBlanc**

Olivier Stocker : **https://polarsensing.net**

[1] Côté, J. F., Widlowski, J. L., Fournier, R. A., & Verstraete, M. M. (2009). The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. Remote Sensing of Environment, 113(5), 1067-1081.
[2] Yong Liu, Rongjun Liu, Jan Pisek, and Jing M. Chen. Separating overstory and understory leaf area in deciduous broadleaf and deciduous broadleaf forests by futso
of MODIS and MISR data. Biogeosciences, 14(5):1309–1326, 2017. Publisher: Copernicus GmbH.
[3] Ewa Burwicz-Galerne, Jan Hackel, Simon Hanocka, Hans Hoyer, Jan Pisek, and Matti Mõttus. Spectral Properties of Coniferous Forests: A Review of In Situ and Laboratory Measurements. Remote Sensing, 16(2):297, 2024. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
[4] Audrey Mercier, Susanna Karlprinz, Aarne Hovi, and Miina Rautiainen. Reflectance spectra of boreal forest floor mosses and lichens and their decrease with wetting. 2, 2023. Publisher: Mendeley Data.
[5] Petri R. Forsström, José Juola, Aarne Hovi, and Miina Rautiainen. Dataset of understory reflectance spectra and fractional cover in a boreal forest area in Finland. 3, 2024. Publisher: Mendeley Data.
[6] Lukas Winiwarter, Alberto Manuel Esmor ́o Pena, Hannah Weiser, Katharina Anders, Jorge Martı ́nez Sánchez, Mariella Searle, Gottfried Mandlburger, et al. Error-in-variables modelling for operational tls-based segmentation of tree stems. ISPRS Journal of Photogrammetry and Remote Sensing, 189:52–68, 2022.