Supplementary Material

A. Simulation Parameter Analysis

To assess the impact of the three simulation parameters, we evaluated each model's performance on the extreme cases within our dataset. We isolated test plots from the lowest 20% and highest 20% of the distribution for tree density, species mixture, and understory vegetation level to analyze their respective effects.

A.1. Tree density

This analysis isolates the effect of tree density on model performance. We compare the results on two subsets of our test data: low-density plots, containing between 5 and 28 trees, and high-density plots, containing between 103 and 128 trees. The detailed results are presented in Table 1.

Impact of Tree Density The most significant finding is the universal degradation in performance as tree density increases. Every model experiences a substantial drop in both Overall Accuracy (OA) and mean Intersection over Union (mIoU) when moving from low-density to high-density plots. The mIoU for DGCNN, for instance, drops by 8.6%, while even the best-performing model, PTv3, sees a drop of over 5%.

For all models, the largest performance drops are seen in the segmentation of tree species and low vegetation. This is likely due to two factors:

- **Crown Overlap**: In high-density plots, the crowns of adjacent trees frequently intertwine, making it extremely difficult for models to delineate the precise boundary between one tree and another.
- **Occlusion**: The denser canopy casts more "shadows" in the point cloud, leading to sparser data for the understory and ground, which complicates their segmentation.

In summary, these results demonstrate that performance in forested environments is directly tied to the forest's structural complexity. By training on a wide spectrum of densities, models can rely on easier samples to learn foundational features before developing the robustness needed to tackle the occlusion and crown overlap in denser plots.

A.2. Species mixture

This analysis isolates the effect of species mixture on model performance. We compare the results on two subsets of our test data: plots where the principal species is Black Spruce, representing the lowest 20% of the Balsam Fir mixture ratio, and plots where the principal species is Balsam Fir, representing the highest 20% of the ratio. The detailed results are presented in Table 2.

The results presented in Table 2 reveal a consistent trend: all benchmarked models perform better on plots dominated by Black Spruce than on those dominated by Balsam Fir. This is demonstrated by a global drop in mIoU of 2.5-4% when Balsam Fir is the principal species.

Table 1. Model performances on plots with the lowest 20% and highest 20% tree densities.

Model	Tree Density	OA	mIoU	IoU$_{Ground}$	IoU$_{AB}$	IoU$_{PM}$	IoU$_{LowVeg}$
MinkUNet	Low	94.3%	90.5%	87.7%	92.2%	94.3%	87.7%
	High	92.7%	84.4%	79.8%	87.9%	89.0%	81.0%
KPConv	Low	94.3%	90.6%	87.7%	93.5%	93.8%	87.4%
	High	92.0%	82.9%	79.1%	87.8%	87.4%	77.3%
DGCNN	Low	92.5%	88.6%	83.3%	93.9%	93.8%	83.5%
	High	89.8%	80.0%	79.0%	84.2%	81.4%	75.5%
Point Mamba	Low	95.9%	90.7%	86.3%	94.9%	95.3%	86.2%
	High	92.8%	83.7%	78.4%	89.7%	89.5%	77.1%
Point Trans. V3	Low	97.1%	94.9%	94.1%	96.4%	96.1%	93.1%
	High	94.3%	89.6%	92.2%	89.0%	88.4%	88.9%

A closer look at IoU scores suggests this performance gap is linked to the structural and spectral characteristics of the two species. While all models successfully segmented the dominant species in each scenario, the overall lower scores in fir-dominant plots indicate they present a more complex segmentation challenge.

We hypothesize this difference is twofold. First, the crown morphology of the trees differs significantly; the typically narrower and more distinct structure of Black Spruce likely results in less canopy overlap. Second, the spectral characteristics of Black Spruce (reflectance of $\sim$0.25) make it easier to differentiate from other forest elements like Balsam Fir ($\sim$0.45) and ground cover (0.44), which have higher reflectance values.

This finding highlights that segmentation accuracy is not just a function of model architecture, but is also significantly influenced by the physical and spectral traits of the target species. As such, our careful use of validated 3D tree models alongside realistic reflectance values brought significant value to our dataset, allowing it to capture these subtle but important challenges.

A.3. Understory vegetation Level

This analysis focuses on the impact of understory vegetation density on model performance. We compare results from two subsets of the test data: "Low Vegetation" plots, which represent the 20% of plots with the sparsest understory, and "High Vegetation" plots, representing the 20% with the densest understory.

Table 3 presents a counter-intuitive trend: all models perform significantly better in high-vegetation scenarios than in low-vegetation scenarios. This contrasts from the effects of tree density and species mixture, where increased complexity led to a universal drop in performance. For instance, Point Transformer v3 mIoU score jumps from 81.6% in low-vegetation plots to 93.4% in high-vegetation plots.

In sparse understory conditions, this class is extremely difficult for all models to identify, with IoU scores ranging from 38.8% (DGCNN) to 58.0% (PTv3). In contrast, when the understory is dense and well-represented in the point cloud, the IoU for this class goes to over 88% for all models except DGCNN.

This suggests that a sparse understory is difficult to de-

Table 2. Model performance on plots with the purest 20% of each species.

Model	Principal Specie	OA	mIoU	IoU$_{Ground}$	IoU$_{AB}$	IoU$_{PM}$	IoU$_{LowVeg}$
MinkUNet	Black Spruce	95.1%	85.5%	85.3%	74.4%	97.1%	85.3%
	Balsam Fir	93.9%	81.9%	83.1%	94.2%	66.0%	84.1%
KPConv	Black Spruce	94.8%	86.4%	84.8%	80.8%	95.8%	84.4%
	Balsam Fir	93.3%	80.7%	82.8%	93.8%	64.4%	81.9%
DGCNN	Black Spruce	93.3%	82.0%	81.4%	70.8%	95.3%	80.5%
	Balsam Fir	93.3%	78.7%	80.1%	95.5%	60.6%	78.5%
Point Mamba	Black Spruce	95.7%	86.7%	83.3%	82.7%	97.3%	83.3%
	Balsam Fir	95.1%	83.7%	82.3%	96.4%	74.6%	81.2%
Point Trans. V3	Black Spruce	97.0%	90.4%	93.3%	79.9%	97.2%	91.4%
	Balsam Fir	96.6%	87.8%	93.2%	96.6%	70.4%	91.1%

Table 3. Model performance on plots with the lowest and highest 20% understory vegetation levels.

Model	Understory Level	OA	mIoU	IoU$_{Ground}$	IoU$_{AB}$	IoU$_{PM}$	IoU$_{LowVeg.}$
MinkUNet	Low	88.4%	72.7%	64.4%	90.2%	91.3%	44.9%
	High	95.3%	91.2%	92.7%	88.6%	90.0%	93.6%
KPConv	Low	90.8%	76.6%	76.7%	90.6%	90.2%	49.1%
	High	94.1%	88.9%	90.5%	88.1%	88.3%	88.9%
DGCNN	Low	84.3%	66.6%	53.9%	87.9%	85.9%	38.8%
	High	93.8%	88.7%	93.4%	85.7%	84.2%	91.5%
Point Mamba	Low	91.2%	77.4%	74.4%	92.8%	92.4%	50.2%
	High	94.2%	89.0%	88.9%	89.1%	90.1%	87.9%
Point Trans. V3	Low	93.3%	81.6%	85.9%	91.6%	91.1%	58.0%
	High	96.5%	93.4%	96.7%	91.0%	90.4%	95.7%

tect. When there are only a few points belonging to low vegetation, models struggle to distinguish them from ground points or noise. However, a dense and voluminous understory provides a much stronger signal that the models can easily learn and identify. The performance on other classes, such as the tree species, remains relatively stable between the two conditions, confirming that the presence of understory vegetation does not significantly complicate the overall segmentation task.

A.4. Summary of Simulation Parameter Analysis

In conclusion, our analysis of the simulation parameters confirms that model performance varies significantly across the extreme cases of our dataset. This sensitivity signals that by deliberately training on a wide spectrum of conditions, our dataset provides a robust training environment. This, in turn, explains the significant 6% mIoU improvement achieved when applying sim-to-real transfer learning.