IEEE IGARSS 2024 July 7-12 ATHENS GREECE

An exploration of properties of point clouds of individual trees extracted from a larger UAV LiDAR survey

Ivan Dubrovin
Skolkovo Institute of Science and Technology

July 10, 2024
IGARSS 2024, Athens

/iod-ine
/in/mr-ivan-dubrovin/

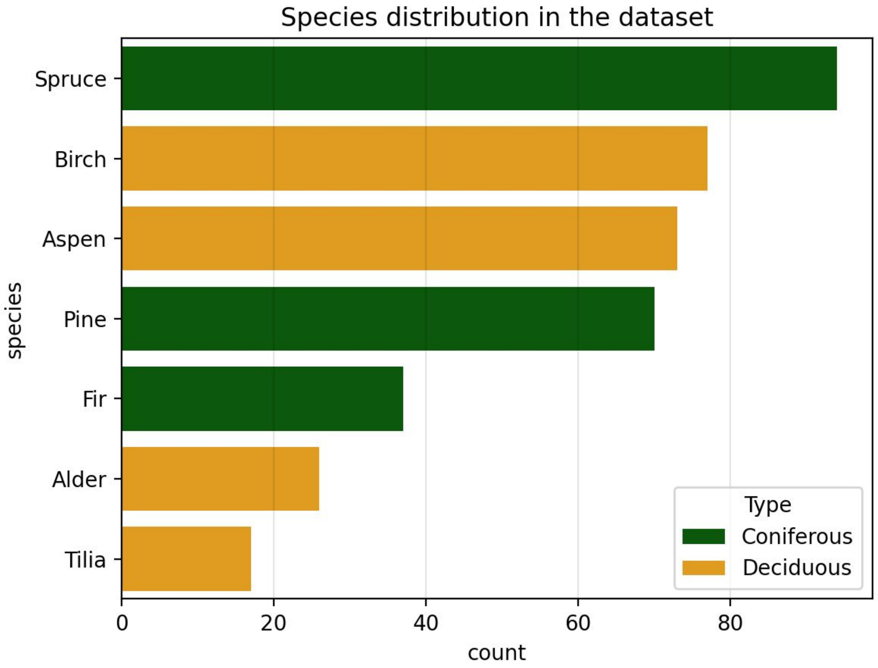
Skoltech

The dataset you will see is freely available to download. The link will be at the end of the presentation.

The dataset: source

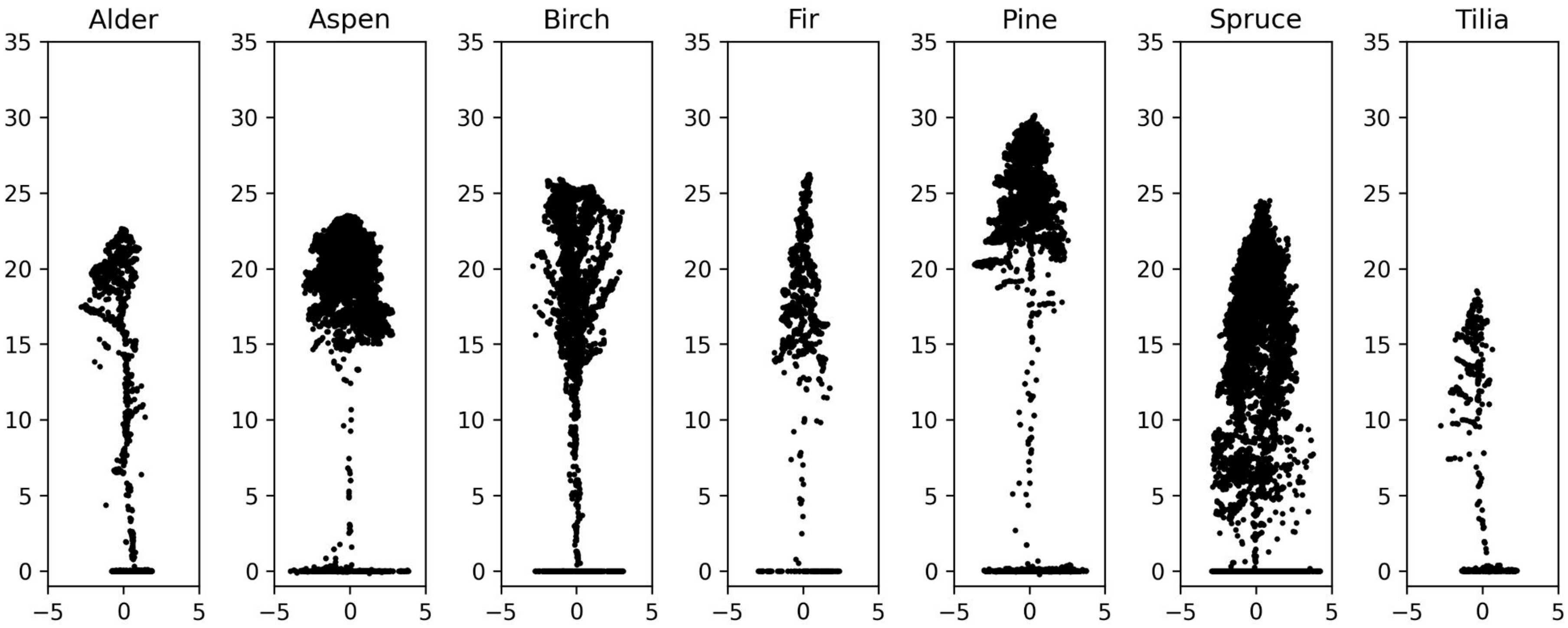
- This project is a part of a larger research effort into using deep learning for modeling dense mixed forests on the scale of individual trees.
- The original source of the data is a large field inventory with overlapping UAV LiDAR and RGB orthophoto.
- The study area is located near Perm, Russia.

The dataset: details

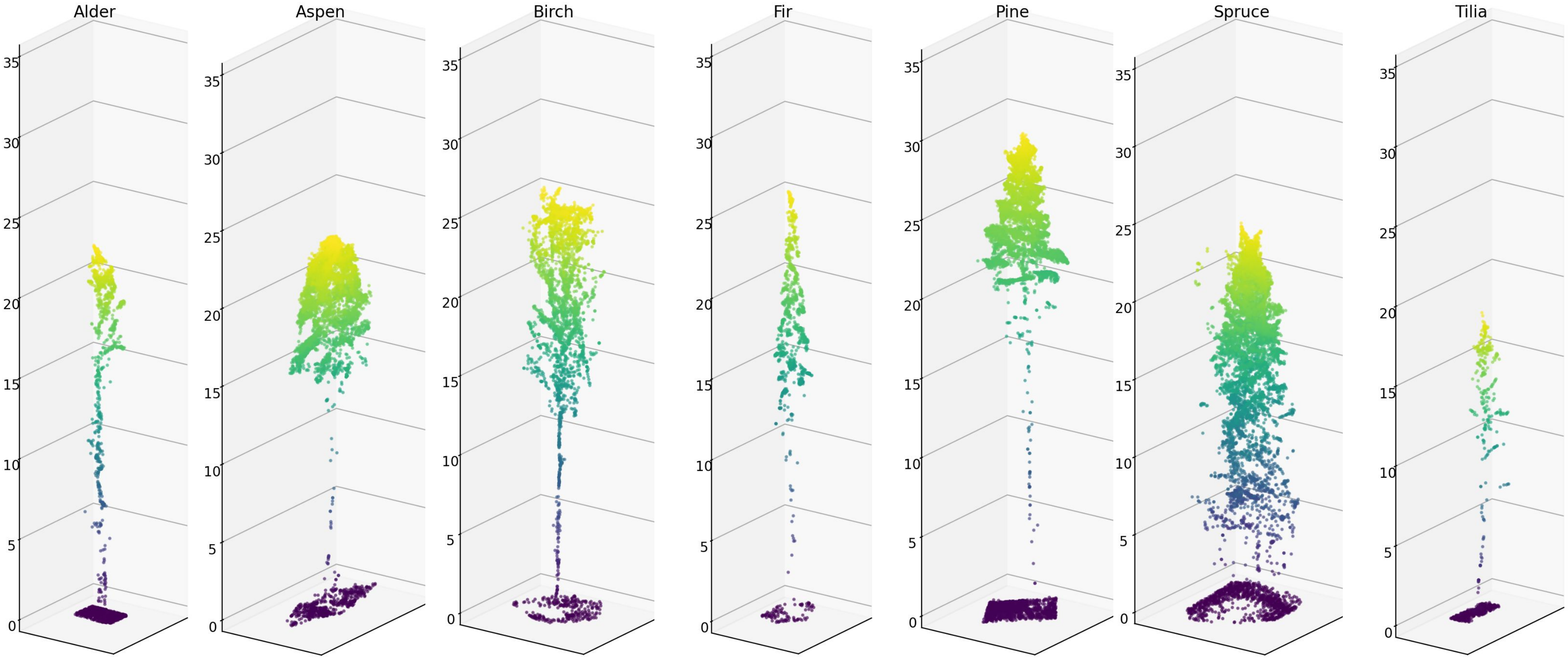


- 394 trees, extracted by hand from a large UAV LiDAR survey.
- Approximately equal number of deciduous and coniferous species.
- Focus on 4 main species: spruce, birch, aspen, and pine.

Cross-sections of individual trees

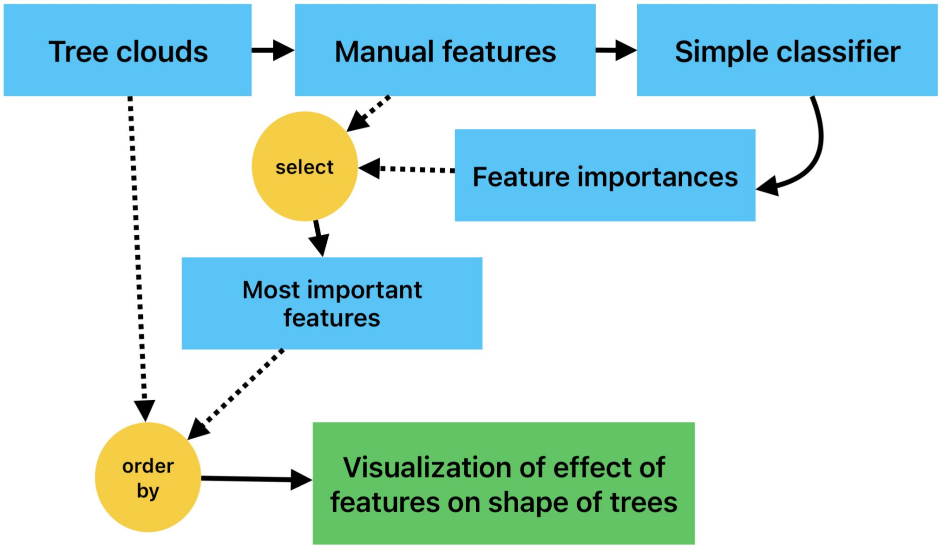


Individual trees in 3D



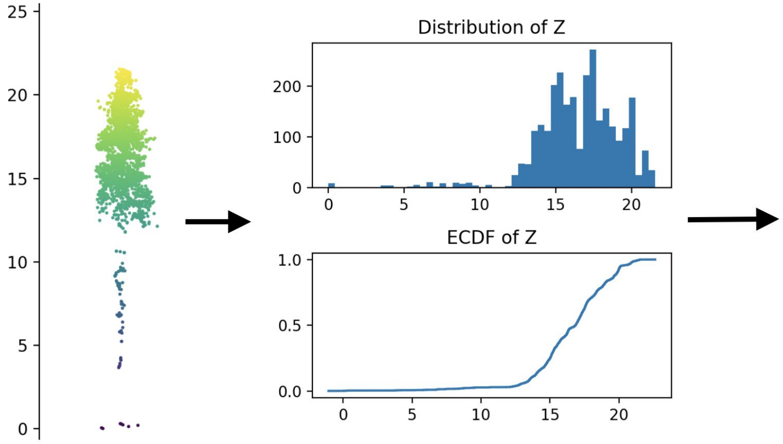
Experiment 1: Shapes of trees <-> ML feature values

- Build intuition into the effect of commonly used classic machine learning features on shapes of individual trees.
- Set up the problem as binary classification: classify each tree into deciduous or coniferous.
- Classification is only required to make sure features have predictive power.



Height distribution features

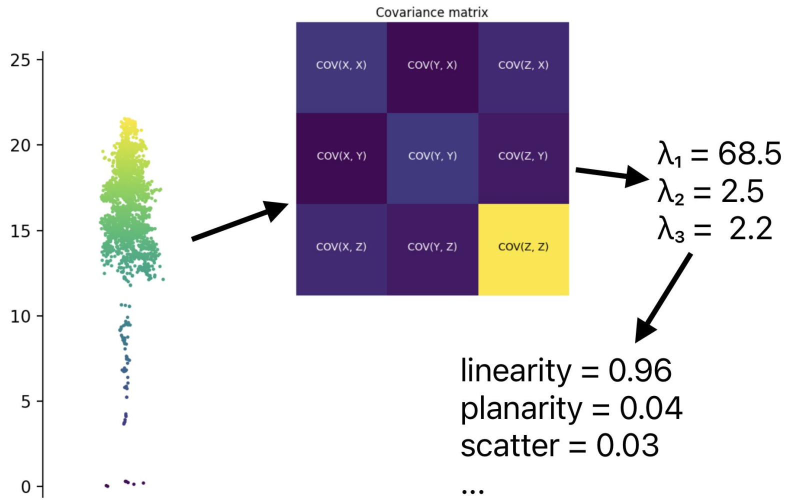
- Maximum height
- Average height
- Standard deviation of height
- Skewness of the height distribution
- Kurtosis of the height distribution
- Entropy of the height distribution
- Proportion of points above the mean
- Proportion of points above 2 m
- Deciles of height
- Cumulative proportion of points at deciles



max = 21.5
mean = 16.5
std = 2.7
...

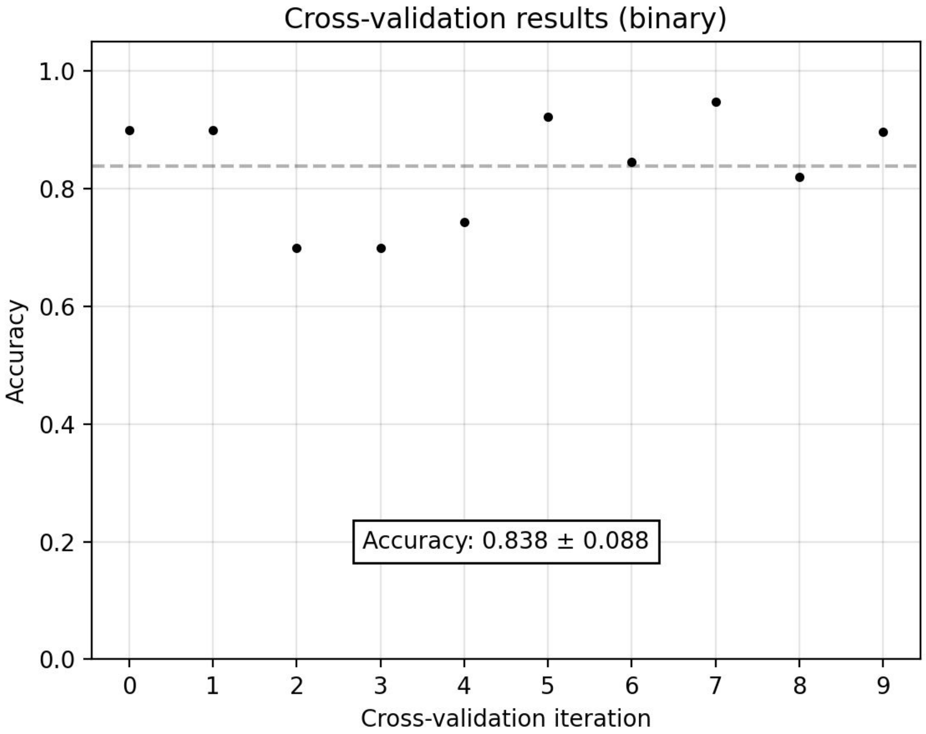
Shape features (based on the eigenvalues of the covariance matrix)

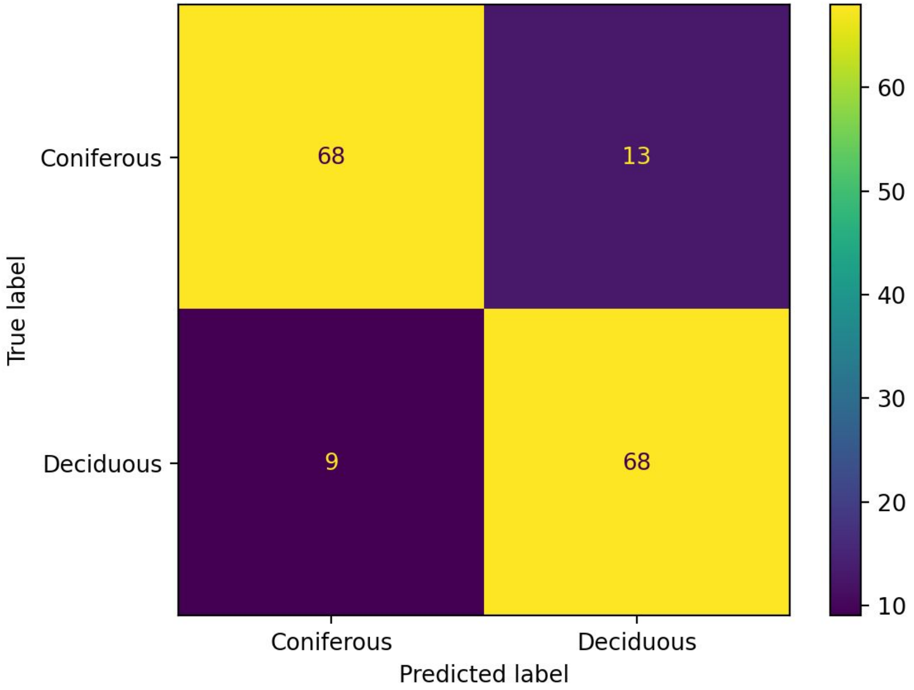
- Linearity: $(\lambda_1 - \lambda_2) / \lambda_1$
- Planarity: $(\lambda_2 - \lambda_3) / \lambda_1$
- Scatter: λ_3 / λ_1
- Omnivariance: $(\lambda_1 * \lambda_2 * \lambda_3)^{1/3}$
- Eigentropy: $-\lambda_1 \log(\lambda_1) - \lambda_2 \log(\lambda_2) - \lambda_3 \log(\lambda_3)$
- Sum of eigenvalues: $\lambda_1 + \lambda_2 + \lambda_3$
- Curvature: $\lambda_3 / (\lambda_1 + \lambda_2 + \lambda_3)$



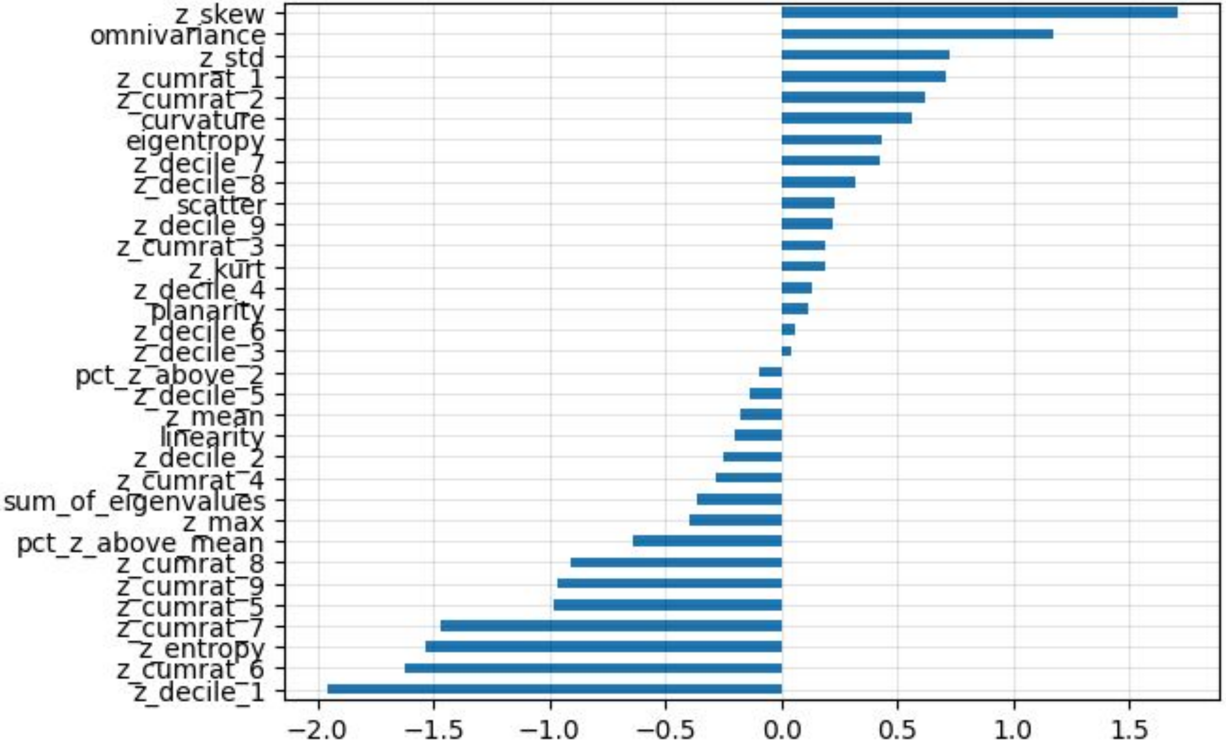
Chris Lucas et al. 2019. “Identification of Linear Vegetation Elements in a Rural Landscape Using LiDAR Point Clouds.”

Logistic regression classification results

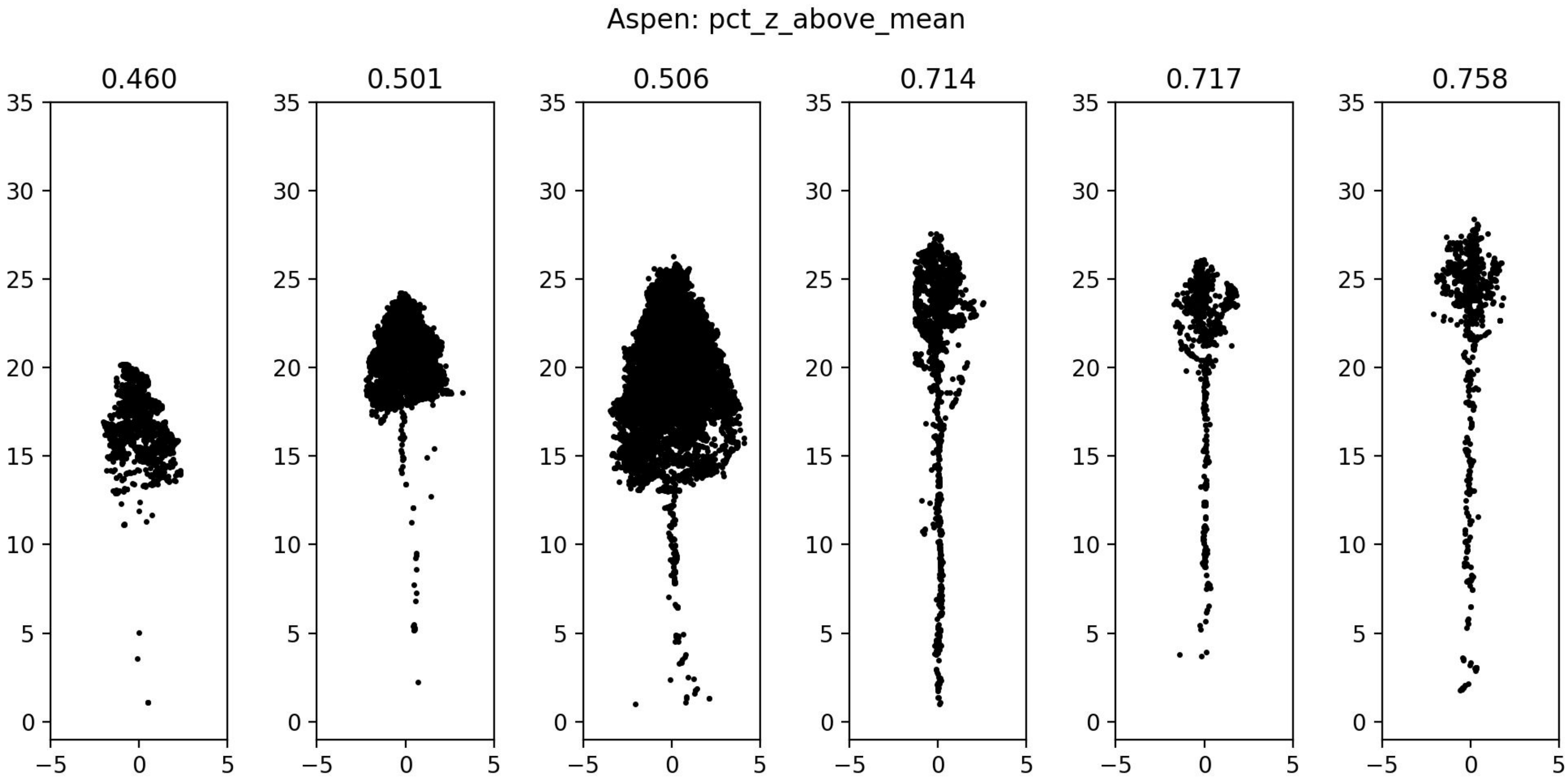




Logistic regression coefficients

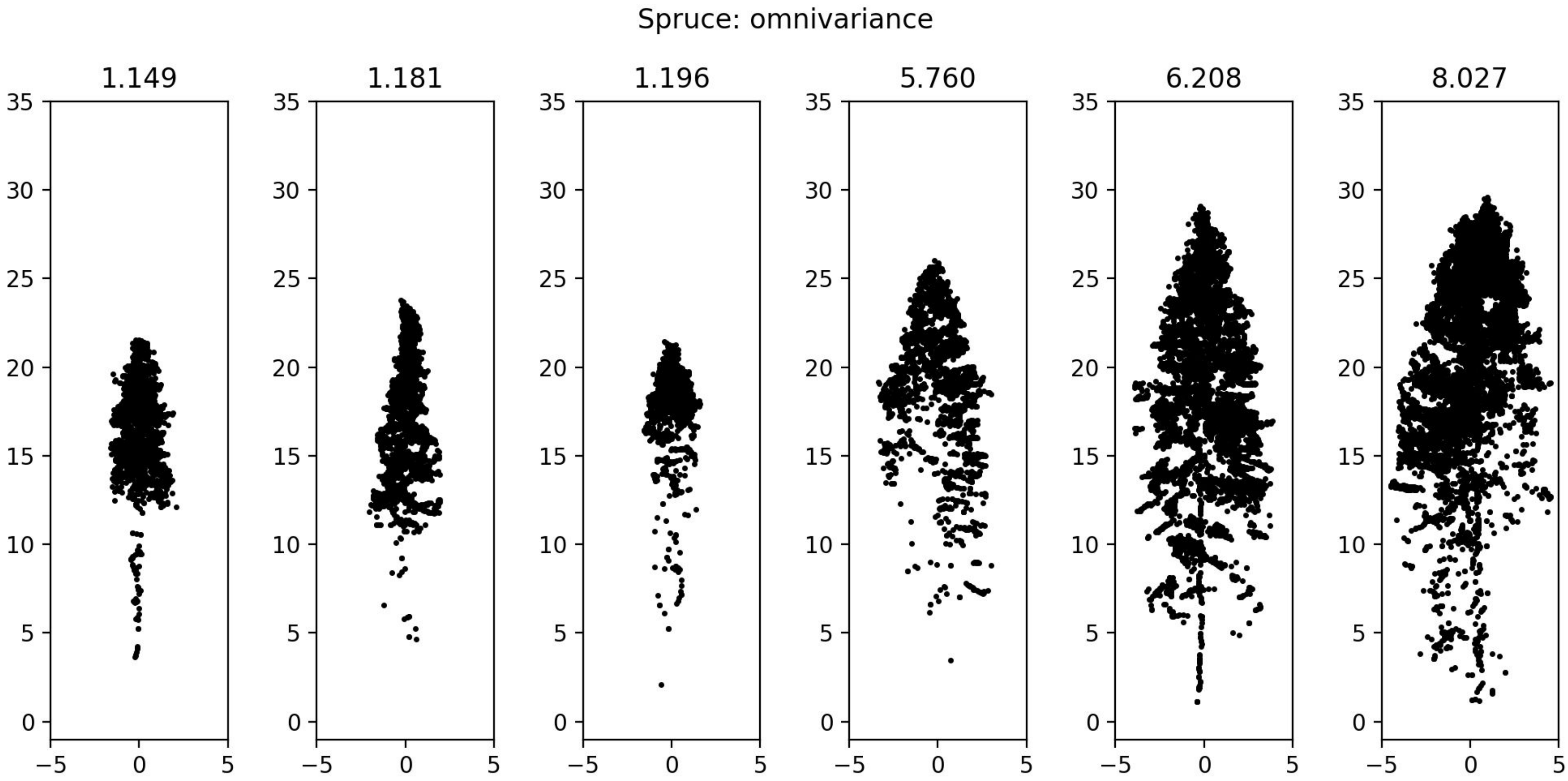


Coefficients of logistic regression are treated as proxy for feature importances to see which features have predictive power.



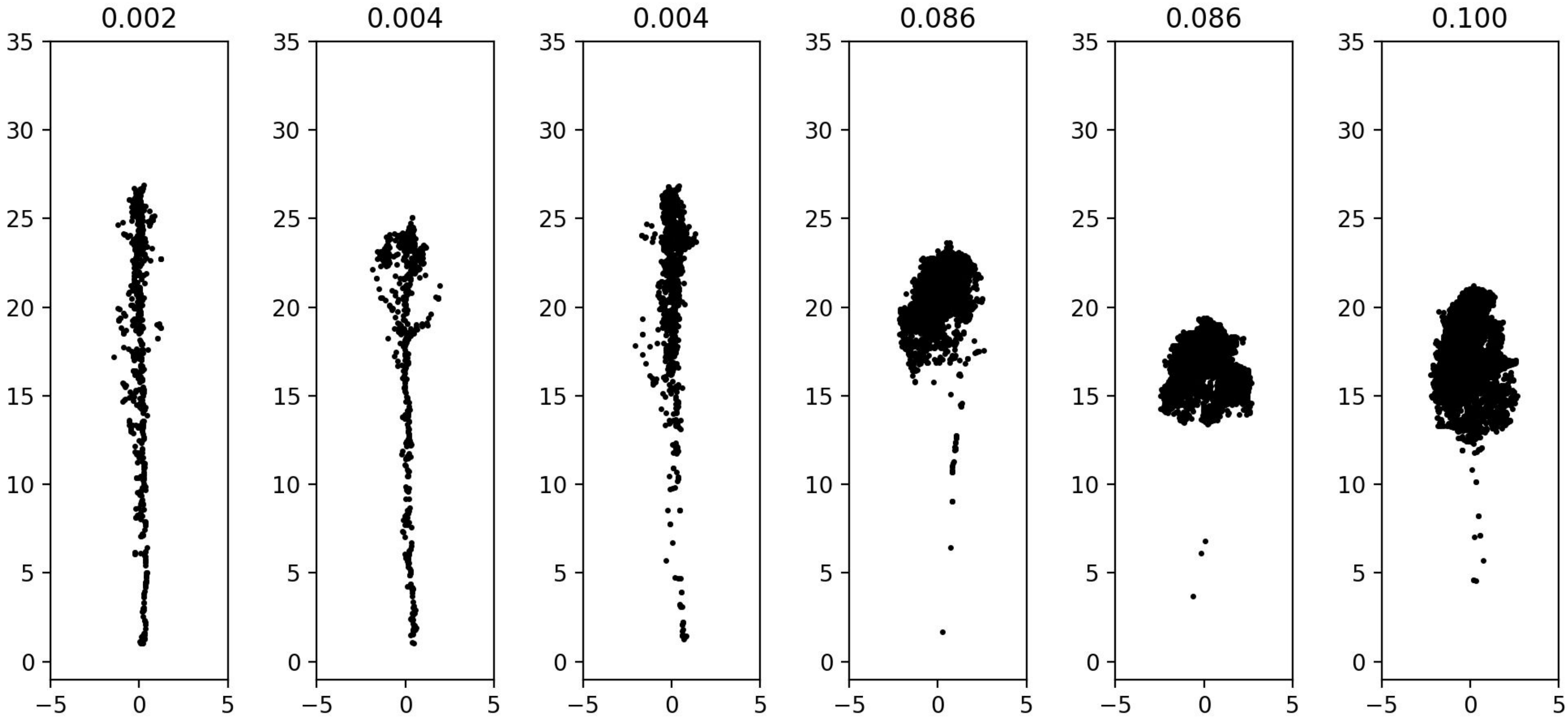
Aspen: pct_z_above_mean
0.460
0.501
0.506
0.714
0.717
0.758
35
30
25
20
15
10
5
0
−5
0
5

Skoltech



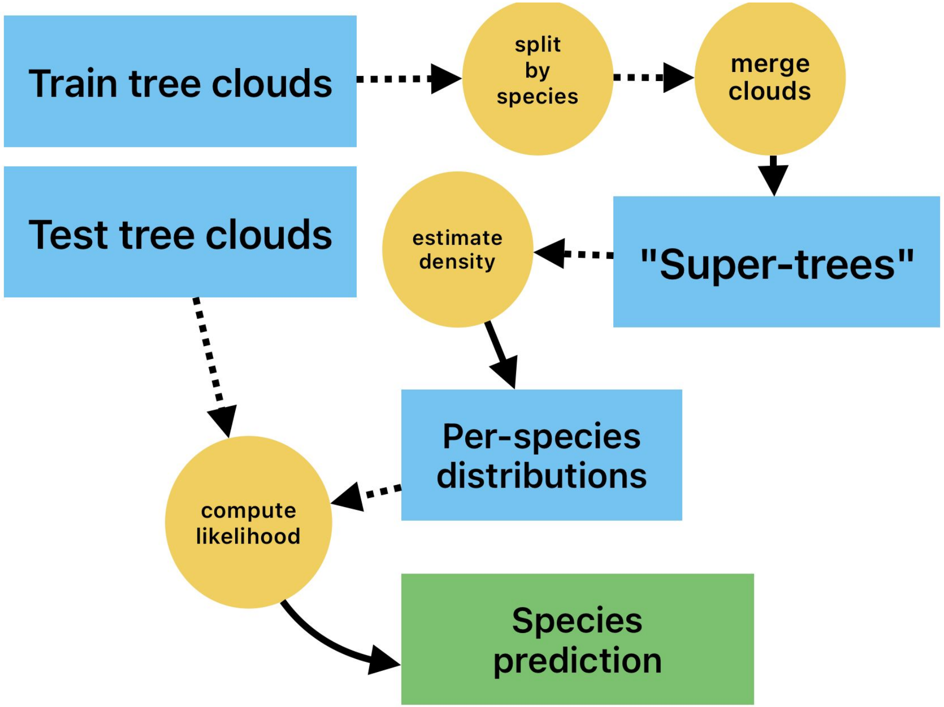
Spruce: omnivariance
1.149
1.181
1.196
5.760
6.208
8.027
35
30
25
20
15
10
5
0
−5
0
5

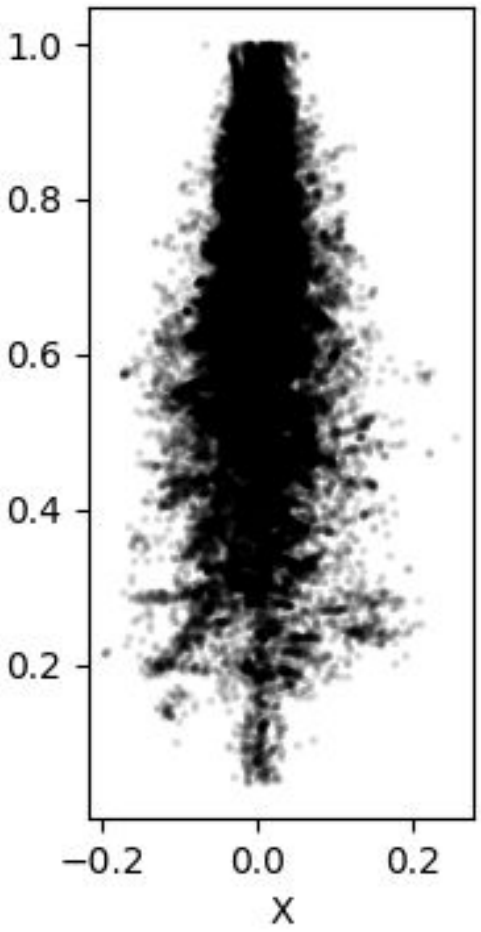
Birch: curvature

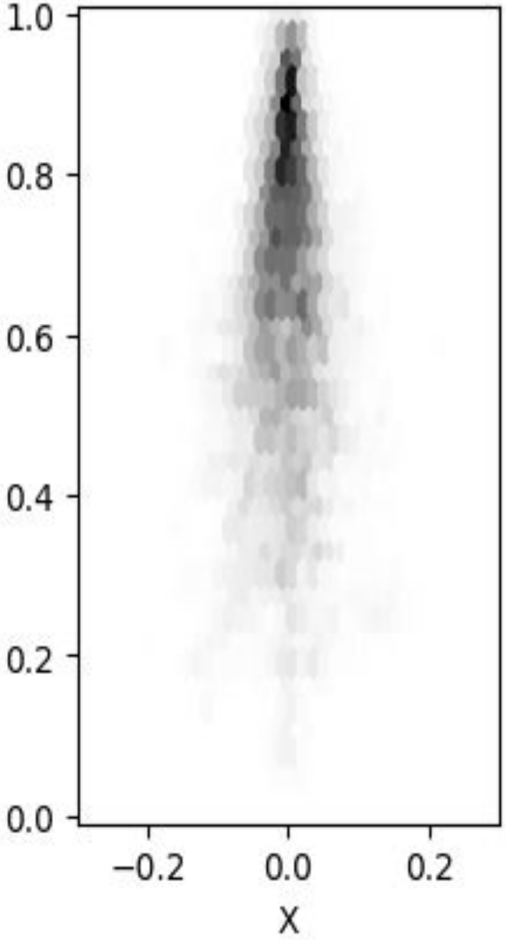


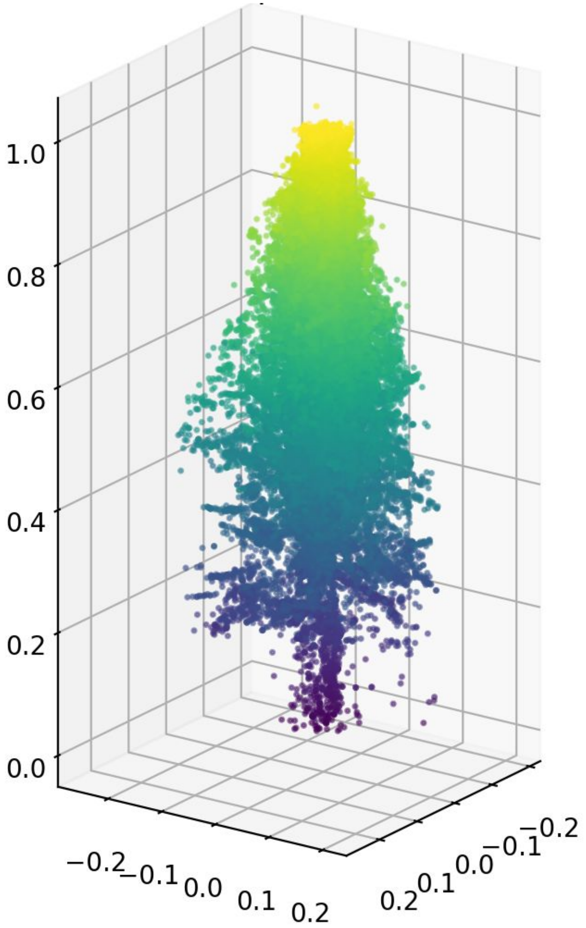
Experiment 2: Tree clouds as 3D distributions

- Scale and merge multiple trees of the same species into a “super-tree”, fit a 3D kernel density estimator for each species.
- For a tree cloud we need to classify, calculate log-likelihoods for each of the per-species densities.
- Classify a tree cloud by assigning the class with the highest likelihood.



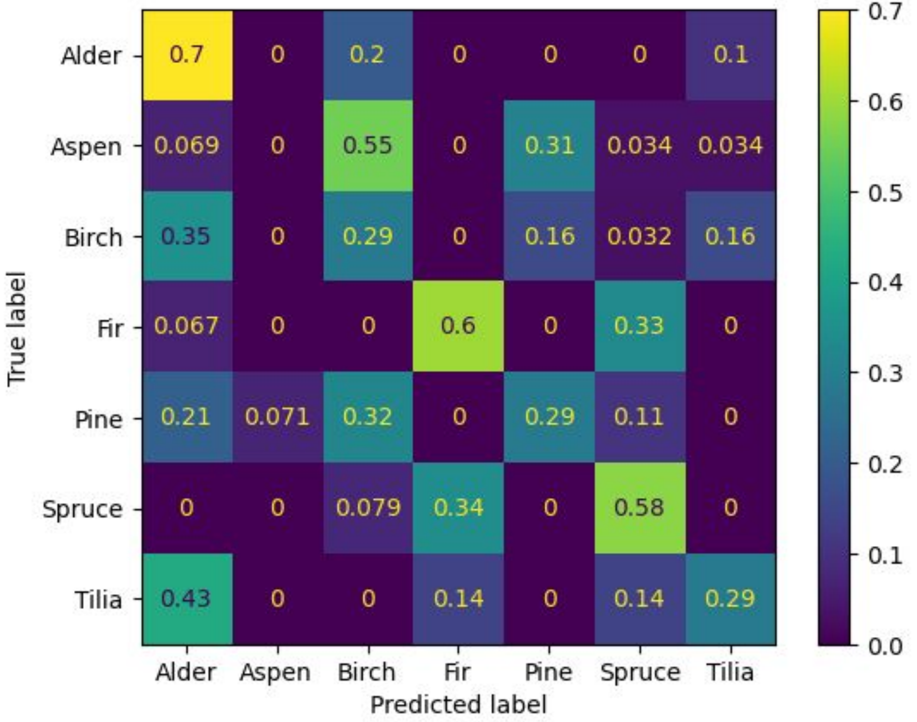

1.0
0.8
0.6
0.4
0.2
−0.2
0.0
0.2
X


1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.2
X


1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.1
0.0
0.1
0.2
0.2
0.1
0.0
−0.1
−0.2

Classification results: multiclass

- Overall accuracy is very low.
- Better distinguishes among coniferous species than deciduous.

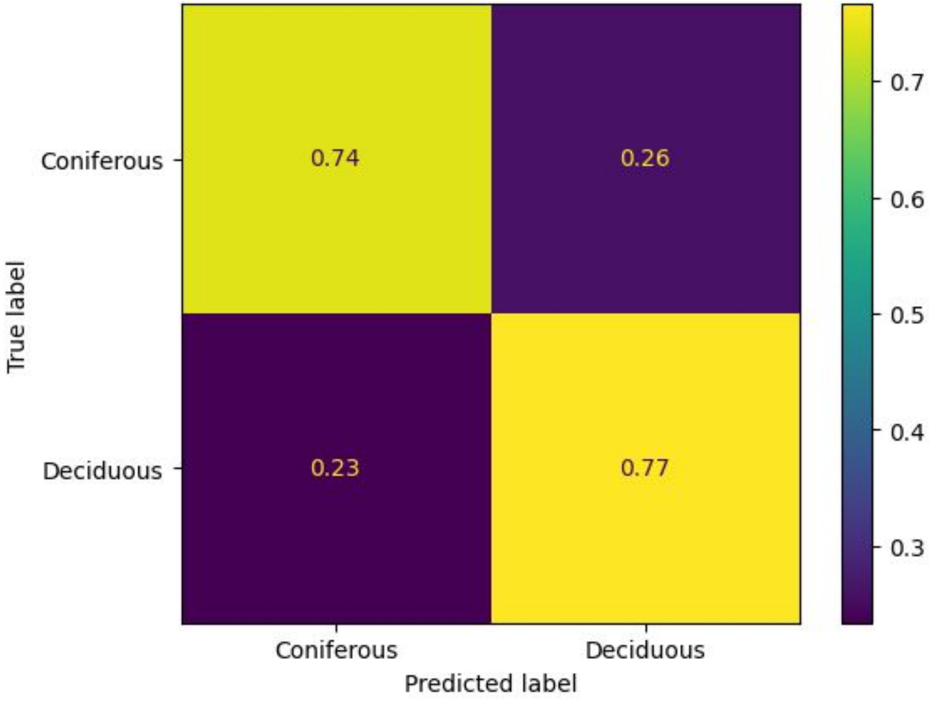


Accuracy: Overall	36.1%
Accuracy: Deciduous	23.4%
Accuracy: Coniferous	48.1%

Classification results: reduced to binary

- This approach distinguishes between deciduous and coniferous trees reasonably well (still worse than unoptimized LR from experiment 1).
- Note: we used the same seven kernel densities, but transformed labels to binary on the fly.

Accuracy: Overall	75.3%
Accuracy: Deciduous	76.6%
Accuracy: Coniferous	74.1%



Final remarks

Experiment 1

- An open dataset, potentially useful in many ways.
- An easy way to visually understand manual features (and any manual metrics that can be derived from a point cloud).

Experiment 2

- Interesting approach, easy to implement, but the results are not good enough.
- The inference is very slow.
- No real increase in result quality with increasing dataset size.

Thank you!

- Follow the link to download the data and accompanying code.
- I will be happy to answer any questions!

kaggle.com/datasets/sentinel3734/uav-point-clouds-of-individual-trees

/iod-ine

/in/mr-ivan-dubrovin/