The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B2-2026
XXV ISPRS Congress 2026 "From Imagery to Understanding", Commission II, 4–11 July 2026, Toronto, Canada

Prediction of Understorey Vegetation using Remote Sensing in Fennoscandian Forests

Ritwika Mukhopadhyay[1*], Ruben Valbuena[1], Inka Bohlin[1]

[1] Dept. of Forest Resource Management, Swedish University of Agricultural Sciences (SLU), 90183 Umeå, Sweden
(firstname.lastname@slu.se)
*corresponding author

Keywords: Airborne laser scanning (ALS), Canopy height model (CHM), Pléiades, Sentinel-2, Tri-stereo image matching (TSIM), Understorey vegetation (USV).

Abstract

Understorey vegetation (USV) contributes to forest structure, nutrient cycling, species diversity, habitat functions, and disturbance processes in Fennoscandian forests. It also provides non-wood forest products such as wild berries. Mapping USV is important for understanding ecosystem functioning and its links to overstorey conditions. Although remote sensing (RS) enables large-scale forest monitoring, its use for USV mapping remains limited because the layer is often obscured by upper-canopy foliage. This study assesses the accuracy of USV cover prediction (i.e., the ground area covered by USV) using multiple RS data sources, identifies key predictors, and evaluates how canopy cover influences model performance. Field data were collected in 2024 from 487 plots in the Krycklan catchment. Sentinel-2 summer and autumn imagery provided spectral reflectance, spectral indices, and grey-level co-occurrence matrix (GLCM) texture variables. Additional texture variables were derived from canopy height models (CHMs) generated using airborne laser scanning (ALS; 1–2 points/m²) and Pléiades tri-stereo image matching (0.5 m; 1.5 points/m²). Beta regression and random forest regression (RFR) models were trained on 70% of plots and validated on 30%. Important predictors included seasonal red-edge differences, greenness-based indices, CHM texture variables, and ALS-based canopy cover. Model performances indicated obstruction due to overstorey canopy cover remains for USV cover prediction. Beta regression with Sentinel-2 data performed slightly better (RMSE = 21.7 m², variance explained = 5%) than RFR. However, best results occurred in low-canopy plots (≤40%) using RFR with Sentinel-2 and Pléiades-derived CHM texture variables (RMSE = 14.6 m², variance explained = 32%).

1. Introduction

The Fennoscandian forests provide essential ecological, cultural, and economic services at high latitudes. One of the components is the understorey vegetation (USV) layer. The USV layer constitutes aboveground biomass, regulates microclimate, nutrient cycling, regeneration dynamics, biodiversity patterns, and overall forest productivity (Nilsson & Wardle, 2005). In Fennoscandian forests, USV influences seedling competition, shapes vertebrate forage and habitat structure, and supports culturally and economically important non-wood forest products, notably bilberry (*Vaccinium myrtillus*), lingonberry (*Vaccinium vitis-idaea*), and crowberry (*Empetrum nigrum*) (Bohlin et al. 2021; Miina et al. 2024). The layer also contributes to surface fuel continuity and fire behaviour, with implications for disturbance regimes and risk management. Understanding the spatial distribution of USV is therefore important to boreal ecosystem assessment and to linking understorey dynamics with overstorey structure.

Remote sensing (RS) has long provided wall-to-wall data for predicting and mapping forest attributes for example, biomass, volume, height, diameter at breast height, and species composition, using area-based approach for model-based inference framework (Nilsson et al., 2017; Ståhl et al., 2011). Yet USV prediction from RS remains challenging as overstorey obstruction restricts top-of-canopy remote sensors' line-of-sight to the forest floor and USV species are typically small, heterogeneous, and patchy, weakening their spectral separability at common satellite spatial resolutions. As a result, most operational mapping pipelines emphasise the top canopy, whereas the USV has received limited attention, with only a handful of empirical attempts to predict USV attributes such as cover or berry yield (e.g., (Bohlin et al. 2021; Miina et al. 2024). Even so, recent advances in multispectral imaging, spaceborne and airborne light detection and ranging (LiDAR) sensor, and high-resolution digital surface models (DSM) now enable indirect characterisation of USV conditions through structural and spectral cues, including canopy height model (CHM) features linked to gap fraction and light penetration.

Because canopy cover modulates light, moisture, and litter inputs, it can also constrain the detectability and prediction accuracy of USV from RS data. Denser, more homogeneous canopies can mask understorey signals, whereas structurally heterogeneous canopies may improve separability by introducing gaps and vertical discontinuities that enhance USV visibility in optical RS data. Consistent with this reasoning, studies of shrub productivity and forest multifunctionality in boreal systems highlight canopy-driven limitations on resources and resilience, while national monitoring programmes underscore both the ecological importance of the understorey and the scarcity of operational, large-area mapping frameworks that include it. Hence, in this study an approach that accounts for canopy structure and seasonal signals when predicting USV cover has been presented. Here, we address these gaps by evaluating spaceborne RS predictors and canopy structural metrics for USV cover mapping in the Krycklan catchment (northern Sweden) using an area-based modelling framework. In this study, we aim to 1) evaluate the accuracy of USV predictions using multiple RS data sources, 2) identify the most important explanatory RS variables, and 3) evaluate the effect of canopy cover to the USV cover prediction accuracies. The USV cover prediction models were developed using spaceborne Sentinel-2 data (spatial resolution of 10 m), canopy height models (CHM) derived from airborne laser scanning (ALS) from the year 2019 and tri-stereo image matching (TSIM) of spaceborne Pléiades data using structure-from-motion algorithm from the year 2023, and field inventory data within an area-based modelling approach. The study contributes new insights into the capabilities and limitations of spaceborne platforms for understorey mapping and explores how horizontal and vertical canopy structural parameters derived from CHM texture analysis influence USV detectability.

2. Materials and Methods

2.1 Study area and field inventory

The study area is the Krycklan catchment area located in Northern Sweden. The catchment covers an area of around 7000 ha in total. The forest comprises of around 63% Scots pine (*Pinus sylvestris*), 26% Norway spruce (*Picea abies*), and deciduous trees such as birch (*Betula pendula* and *Betula pubescens*) have scattered occurrence. The understorey layer is dominated by ericaceous shrubs, namely, bilberry (*Vaccinium myrtillus*) and lingonberry (*Vaccinium vitis-idaea*), and dense moss (Larson et al. 2024). The field inventory plots in Krycklan are in total 556 plots (10 m radius) with a nominal square grid with 350 m spacing was laid (located in forest- by assuming 10% of the plots to fall in non-forest land) (Swedish National Forest Inventory and Swedish Soil Inventory 2021). Out of the 556 field inventory plots, ground measurements from only 487 plots have been used in this study that were measured during the year 2024, as presented in Figure 1. At each 10 m radius plot the area considered for the USV measurements is a circular plot with a radius of 5.64 m (100 m^2). The shrub species occurrence and coverage are estimated by trained field personnel at each field inventory plot. The cover is estimated vertically from the projection of the plants on the ground surface by the shadow the plants would cast on the ground if they were illuminated from above with parallel light beams (Swedish National Forest Inventory and Swedish Soil Inventory 2021). The total cover of the USV layer is measured along with the cover of individual species for the 5.64 m radius plots as percentiles and then presented as estimates in m^2 for the entire 10 m radius plots. In this study, we have only focused on the total cover of the USV layer.

2.2 RS data processing

Spaceborne RS data: Sentinel-2 multispectral time-series data of 10 m spatial resolution were acquired and processed in open-access Google earth engine platform for the year 2024. The Sentinel-2 images were acquired for two separate seasons, namely – summer (June-August) and autumn (September-December). Total number of cloud free images acquired for summer and autumn were 12 and 22 images, respectively. Geometric median composites were derived for the cloud free images for the two separate seasons. Geometric median image composites have been observed to preserve the correlation among the individual spectral bands when they are aggregated over several acquisitions (Roberts, Mueller, and Mcintyre 2017). Several spectral indices were derived for each season separately from the Senitnel-2 geometric median image composites, where spectral indices reflect distinct ecological, plant physiology phenomenon, and structural dependencies within the forest stands. For example, the Normalized Green-Red Index (NGRDI), Visible Atmospherically Resistant Index (VARI), Excess Green Index (EXG), Normalized Green-Blue Difference Index (NGBDI), Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), Green-Red Ratio Index (GRRI), Enhanced Vegetation Index (EVI), Green Normalized Difference Vegetation Index (GNDVI), Red-Edge Normalized Difference Vegetation Index (RENDVI), Modified Soil-Adjusted Vegetation Index (MSAVI), and Chlorophyll Index Red-Edge (CIRE). Also, differences between the derived spectral indices for summer and autumn seasons were computed. Out of all the derived spectral indices, NGRDI highlights the green vegetation and is derived using the green and red visible bands, as represented in Eq. (1) and the visible atmospherically resistant index (VARI) also highlights green vegetation by reducing the sensitivity of the atmospheric effects and is also derived using the visible spectral bands – green, red, and blue, as represented in Eq. (2). Grey-level co-occurrence matrix (GLCM) texture values were extracted from the spectral reflectance and the spectral indices derived from the geometric median image composites for a 3×3 window size.

$$\mathrm{NGRDI} = \frac{G-R}{G+B}, \tag{1}$$

$$\mathrm{VARI} = \frac{G-R}{G+R-B}, \tag{2}$$

where R, G, B = Red, green, and blue visible spectral bands, respectively, from Sentinel-2 RS data.

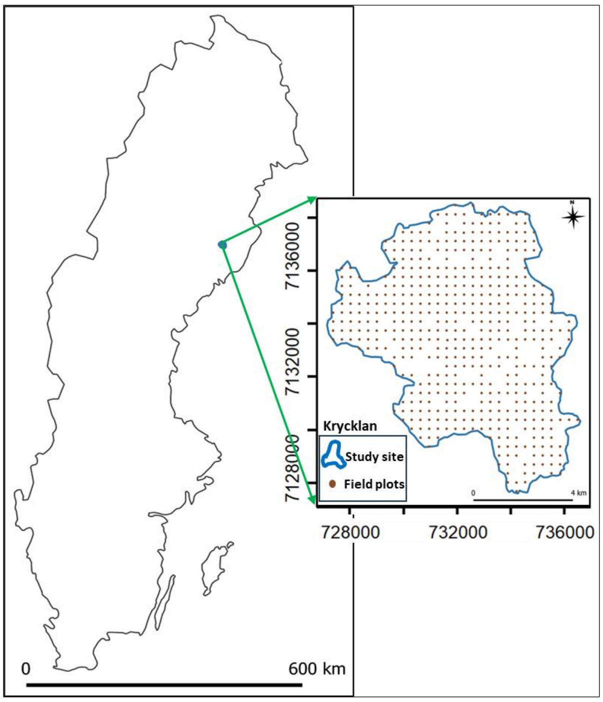


Figure 1. Shows the study site – Krycklan, Northern Sweden and the layout of the systematic field inventory plots over the entire Krycklan Catchment area (Larson et al. 2024).

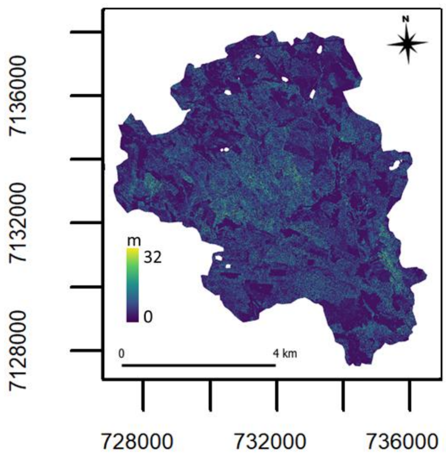


Figure 2. Pléiades TSIM derived CHM raster over the Krycklan Catchment area.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-501-2026 | © Author(s) 2026. CC BY 4.0 License.

Pléiades multispectral tri-stereo images were acquired for the year 2023 over the study area. Structure-from-motion algorithm was implemented to re-construct a 3-dimensional (3D) point cloud of 1.5 points/m^2 from tri-stereo image matching (TSIM) and further derive a canopy height model (CHM) raster of 0.5 m spatial resolution for the entire study area using *Agisoft* commercial tool, as presented in Figure 2. GLCM texture values, such as, mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation, were extracted from the Pléiades TSIM-based CHM raster for a similar 3×3 window size (Fassnacht et al. 2021).

The explanatory variables for the USV models were extracted from the Sentinel-2 image composites, and the ALS- and Pléiades-derived CHM rasters over the field plot areas using R tools. The digital number (DN) values were extracted as mean through an area-based approach and as point estimates for the plot centres for the individual spectral bands, the spectral indices for both seasons – summer and autumn, and the ALS- and Pléiades-derived CHM rasters.

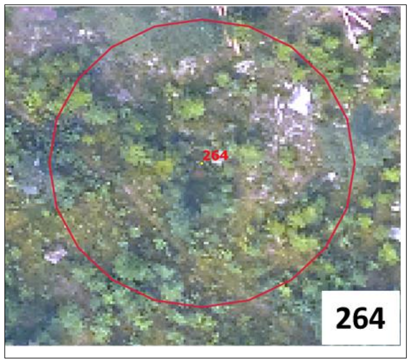
2.2.1 Airborne RS data: ALS campaigns have been conducted mainly to develop a nationwide digital elevation model by the Swedish National Mapping Agency during the years 2009-2017 (Nilsson et al., 2017). Forest attribute maps have been generated as well from the acquired ALS data for entire Sweden, for example, volume, basal area, average height, average diameter, and aboveground biomass (Nilsson et al., 2017). After 2017, the second scanning campaign has been conducted between the years 2018-2025. For this study, the ALS point clouds (1-2 points/m^2) from the second scanning were extracted for the study area and processed using *lastools* to derive a CHM raster for the study area. A canopy closeness metric also mentioned as the vegetation ratio (vr) is a proxy variable for top layer canopy coverage was also derived from the ALS point clouds (Nilsson et al., 2017) and as expressed in Eq. (3),

$$vr = \frac{\text{percentage of first returns above 1.5 m}}{\text{total number of first returns}}, \tag{3}$$

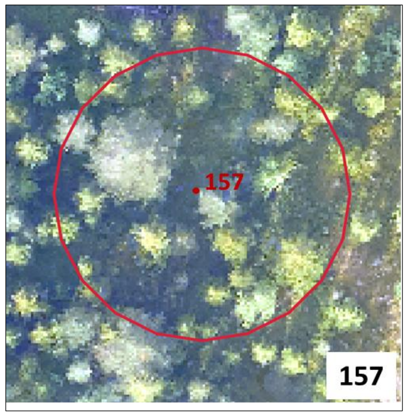
2.3 Statistical modelling

The field measured cover of the overall USV layer per plot were scaled between 0 and 1. Beta regression models and random forest regression (RFR) were fitted with the field measured cover of the USV layer as the response variable and the spectral and canopy cover information derived from the RS data were used as the explanatory variables. Out of all the extracted data from the individual spectral bands, the spectral indices for both seasons – summer and autumn, and the ALS- and Pléiades-derived CHM texture values the variables presented in Eq. (4) and (5) were identified as important and selected as explanatory variables for the final models through iterative forward and backward selection approach using R tools. Along with the variable selection approach, manual correlation between field measured USV cover and all variables were also checked.

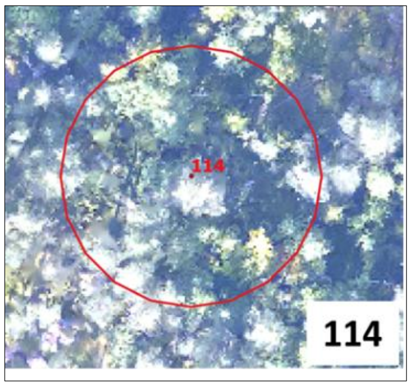
The USV cover prediction model was developed first with using only the spectral reflectance and spectral indices information extracted from Sentinel-2 image composites as explanatory variables as represented in Eq. (4). The developed model (Eq. (5)) was further expanded with GLCM texture values computed over the CHM derived from ALS and Pléiades data, as represented in Eq. (5). For both, the Beta regression models and the RFR, 70% of the field reference data, i.e., 342 plots were used for training and the rest 30%, i.e., 145 plots were used for validation. The predicted USV cover values were rescaled back to the original field measured USV cover range of values between 0 to 100 m^2. The validation dataset was categorized into 3 canopy cover classes: low (40% and below; 28 plots), medium (40-70%; 60 plots), and high (70% and above; 57 plots) and examples presented in Figure 3.



(a)



(b)



(c)

Figure 3. Represents the three plots for the different canopy cover classes: (a) Plot 264 - low (40% and below); (b) Plot 157 - medium (40-70%); (c) Plot 114 - high (70% and above).

The USV models:

$$\widehat{U}_{cover} = \beta_0 + \beta_1 \mathrm{NGRDI}_A + \beta_2 \mathrm{VARI}_A + \beta_3 \overline{\mathrm{RE}}_{S-A}, \tag{4}$$

$$\widehat{U}_{cover} = \beta_0 + \beta_1 \mathrm{NGRDI}_A + \beta_2 \mathrm{VARI}_A + \beta_3 \overline{\mathrm{RE}}_{A-S} + \beta_4 vr + \beta_5 \mathrm{CHM}_{Hom} + \beta_6 \mathrm{CHM}_{Con} + \beta_7 \mathrm{CHM}_{Diss}, \tag{5}$$

where $\widehat{U}_{cover}$ = USV cover scaled between 0 and 1,
NGRDI = plot-centre value of normalized green-red difference index,

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-501-2026 | © Author(s) 2026. CC BY 4.0 License.

VARI = plot-centre value of visible atmospherically resistant index,
$\overline{\text{RE}}$ = mean value of red-edge band of Sentinel-2
Suffixes S and A = denote summer and autumn season, respectively,
vr = Vegetation ratio used here as the proxy canopy closeness metric derived from ALS point clouds as a ratio of percentage of first returns above 1.5 m to the total number of first returns,
CHM_x = CHM texture metrics derived from ALS and Pléiades data,
β_x = model coefficients.

3. Results

3.1 USV cover prediction model coefficients and explanatory variables

The explanatory variables for predicting USV cover in Eq. (4) and (5) are presented in Table 1, together with their coefficients and associated statistical significance. Across all model configurations, the Beta regression results indicate consistent and significant contributions from both spectral and structural predictors for estimating USV cover. In the model relying solely on Sentinel-2 spectral variables (Eq. 4), coefficients β_1, β_2, and β_3 for variables NGRDI, VARI, and red-edge (RE) band, respectively, were statistically significant, indicating that the selected spectral reflectance information alone carries explanatory power for variation in USV cover. When the selected ALS-derived CHM texture metrics were included (Eq. 5), additional coefficients ($\beta_4 - \beta_7$) were also statistically significant, suggesting that canopy structure variability provides complementary information beyond coarse-resolution spectral data. A similar pattern was observed for the Pléiades-derived CHM model, in which all structural coefficients were statistically significant, suggesting that canopy structure from high-resolution Pléiades contributes to explaining USV cover.

Since in this study the indices were also computed for two different seasons separately, for example, summer and autumn image composites, the indices such as NGRDI and VARI from the autumn composites were statistically significant in explaining the USV cover in Eq. (4) and (5). NGRDI showed a positive and statistically significant relationship with USV cover across all model configurations. In contrast, VARI and difference of mean RE reflectance between summer and autumn months ($\overline{\text{RE}}_{S-A}$) exhibited a negative and statistically significant coefficient, indicating that higher VARI and $\overline{\text{RE}}_{S-A}$ values correspond to lower USV cover, as seen from Table 1. The statistically significant texture variables derived from the ALS- and Pléiades-derived CHM in Eq. (5) were homogeneity, contrast, and dissimilarity.

3.2 Model performance across canopy cover classes

3.2.1 Model performance on training data

Beta regression: Across the training dataset, Beta regression models showed highest accuracy in low-canopy conditions (≤40% canopy cover) (Figure 4-5 and Table 2 (Appendix)). Using Sentinel-2 spectral information alone, the model achieved a root mean squared error (RMSE) of 17.3 m² with the highest variance explained (14%) in low canopy cover plots, compared with higher errors and lower explanatory power in medium (5%), high (8%), and overall (9%) canopy cover classes. Adding ALS-derived CHM texture variables produced only a marginal improvement in overall variance explained (from 9% to 11%) and improved performance in high canopy cover plots (8% to 11%). However, the benefit in low canopy cover areas decreased (variance explained dropped from 14% to 8%, and RMSE increased from 17.3 m² to 18.4 m²).

USV Beta regression	Coeff.s	Values
USV~S-2 spectral information (Eq. (4))	β_0	-0.64***
	β_1	9.85**
	β_2	-4.12’
	β_3	-0.01×10^{-1}***
USV~S-2 spectral information + ALS CHM texture variables (Eq. (5))	β_0	1.09×10^{3}*
	β_1	9.50**
	β_2	-3.91’
	β_3	-6.17×10^{-4}***
	β_4	-1.02×10^{-2}**
	β_5	-1.08×10^{3}*
	β_6	1.08×10^{2}*
	β_7	-6.51×10^{2}*
USV~S-2 spectral information + PL CHM texture variables (Eq. (5))	β_0	1.55×10^{3}***
	β_1	1.04×10^{1}**
	β_2	-3.93’
	β_3	-8.35×10^{-4}***
	β_4	-6.58×10^{-3}**
	β_5	-1.55×10^{3}***
	β_6	1.56×10^{2}***
	β_7	-9.29×10^{2}***

Table 1. Model coefficients for Eq. (4) and (5) for Beta regression, where, S-2 denotes Sentinel-2, PL denotes Pléiades, β_x are the model coefficients, and *Asterisks* denote statistical significance: *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$, ’ = $p < 0.1$.

The integration of Pléiades-derived CHM texture variables led to inconsistent effects. While RMSE in low canopy cover plots remained comparable (18.2 m²), the variance explained was moderate (9%). Model performance did not improve over the Sentinel-2-only model, and errors increased in medium canopy cover plots (RMSE = 23.3 m²). Overall, Beta regression using training data explained 5–14% of variance depending on canopy cover class and model type, with performance consistently better in low canopy cover plots.

RFR: RFR outperformed Beta regression during training, especially when CHM texture variables were included (Figure 6-7 and Table 3 (Appendix)). With Sentinel-2 spectral data alone, RFR already showed improved accuracy in low-canopy plots (RMSE = 18.3 m²; variance explained = 7%), though performance remained poor in medium and high canopy cover classes with explained variance ≤ 3%. The inclusion of ALS-derived CHM texture variables provided only minor improvements. Variance explained decreased in low canopy cover plots (7% to 3%), and RMSE remained similar (18.6 m²). Medium canopy cover plots showed no improvement (0% variance explained), indicating limited additional structural information captured by ALS CHM for canopy cover ≤ 70%.

In contrast, the model incorporating Pléiades-derived CHM texture variables produced the highest training performance among all configurations. Low-canopy plots showed the most improvement, with RMSE being reduced to 17.3 m² and variance explained rising to 14%. This demonstrates that the finer spatial detail captured by Pléiades-derived CHM texture metrics effectively supports the non-linear splitting of RFR. Across the full training dataset, RFR explained 1–14% of the variance, with a better performance consistently in low canopy cover plots and when Pléiades-derived CHM variables were included.

3.2.2 Model performance on validation data

Beta regression: Validation results showed patterns similar to those observed in training, with Beta regression performing better in low canopy cover plots and weak in medium, overall, and high canopy cover classes, respectively (Figure 4-5 and Table 4 (Appendix)). Using Sentinel-2 data alone, the model yielded the lowest RMSE of 16.3 m^2 and the highest variance explained (20%) in low canopy cover areas. Medium, overall, and high canopy cover classes showed higher errors (24.9 m^2, 21.7 m^2, and 20.5 m^2, respectively) and substantially lower variance explained (≤4%).

Adding ALS-derived CHM texture variables did not enhance validation performance. RMSE increased in low canopy cover plots (16.3 to 18.2 m^2), and variance explained decreased up to (20% to 5%). Similar patterns were observed for all canopy cover classes, indicating that ALS-based canopy structural texture metrics contributed potentially to model overfitting during training.

RFR: RFR showed improved generalization compared with Beta regression, particularly in low canopy cover plots (Figure 6-7 and Table 5 (Appendix)). With Sentinel-2 alone, RFR achieved variance explained values up to 10% in low canopy cover stands, with an RMSE of 17.2 m^2, outperforming Beta regression. When ALS-derived CHM texture variables were added, performance was observed to improve in low canopy cover plots (RMSE = 15.9 m^2; variance explained = 21%). Medium canopy cover plots also showed a gain in variance explained (0% to 4%), although high canopy cover plots had no improvement.

4. Discussions

4.1 Interpretation of spectral variables and influence of canopy cover on USV cover prediction

The spectral indices derived from Sentinel-2 – NGRDI, VARI, and the seasonal red-edge difference ($\overline{\mathrm{RE}}_{S-A}$), each contributed uniquely to characterizing USV cover, reflecting distinct ecological, plant physiology phenomenon, and structural dependencies within the forest stands. The positive association between NGRDI from autumn imagery and USV cover indicates that greenness-based indices are effective for detecting understorey vegetation wherever canopy openness allows direct observation of the forest-floor vegetation layer. Herbaceous and shrub-layer species typically maintain strong green reflectance during their active growth periods, enabling indices such as NGRDI to capture fine-scale variation in USV density. Higher NGRDI values in the plots are therefore consistent with higher USV cover because the index amplifies the contrast between photosynthetically active vegetation and background non-vegetated surfaces. In contrast, VARI and $\overline{\mathrm{RE}}_{S-A}$ show negative relation with USV cover. VARI is strongly influenced by reflectance from the uppermost sunlit foliage, such that increases in canopy greenness or canopy closure elevate VARI values while simultaneously reducing the visibility of the vegetation beneath the canopy. The negative relationship between VARI and USV cover thus reflects the strong influence of overstorey phenology and structural dominance, which can obscure understorey vegetation even when USV is present.

A similar mechanism explains the negative coefficient of $\overline{\mathrm{RE}}_{S-A}$. The red-edge region of the Sentinel-2 spectrum is sensitive to chlorophyll content and leaf structural properties. Seasonal

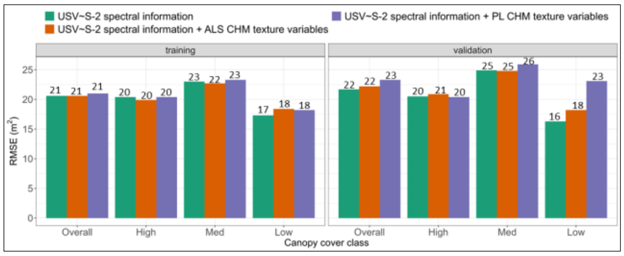


Figure 4. Statistical summary representing the RMSE values (m^2) (rounded to the nearest whole number and presented on each bar and units applicable from the Y-axis) for training and validation data for Eq. (4) and (5) fitted and tested for Beta regression.

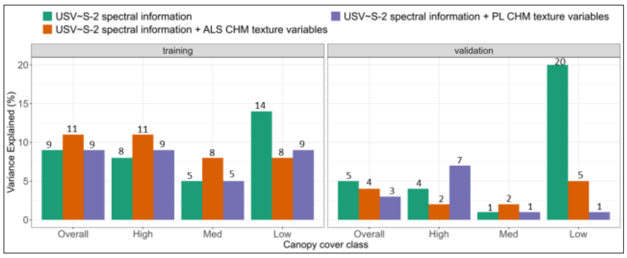


Figure 5. Statistical summary representing the Variance explained (%) (rounded to the nearest whole number and presented on each bar and units applicable from the Y-axis) for training and validation data for Eq. (4) and (5) fitted and tested for Beta regression.

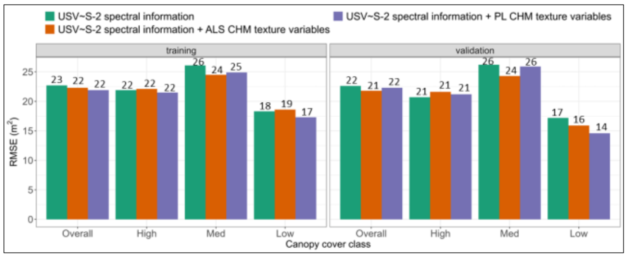


Figure 6. Statistical summary representing the RMSE values (m^2) (rounded to the nearest whole number and presented on each bar and units applicable from the Y-axis) for training and validation data for Eq. (4) and (5) fitted and tested for random forest regression (RFR).

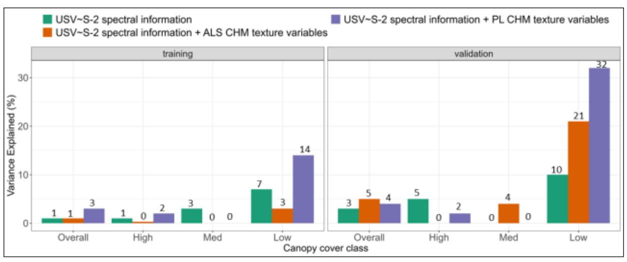


Figure 7. Statistical summary representing the Variance explained (%) (rounded to the nearest whole number and presented on each bar and units applicable from the Y-axis) for training and validation data for Eq. (4) and (5) fitted and tested for random forest regression (RFR).

changes in red-edge reflectance are therefore primarily driven by the overstorey canopy, which undergoes summer-to-autumn shifts in pigment concentration and internal leaf structure which might be mainly driven by the presence of scattered deciduous trees. Plots exhibiting relation to $\overline{\mathrm{RE}}_{S-A}$ correspond to stands with phenologically dynamic overstoreys. Consequently, higher $\overline{\mathrm{RE}}_{S-A}$ values do not necessarily indicate lower USV cover because of ecological processes but can rather indicate reduced spectral visibility of the understorey due to canopy dominance. These negative associations likely represent a remote-sensing occlusion effect rather than a genuine decline in USV abundance.

Overall, the spectral predictors describe complementary aspects of forest structure: NGRDI captures USV greenness, VARI represents reflectance from the upper canopy layer, and $\overline{\mathrm{RE}}_{S-A}$ reflects overstorey structural and phenological dominance. Their combined behaviour indicates that accurate USV prediction depends both on detecting the presence of USV and on quantifying the canopy conditions that limit its visibility. This interpretation is supported by model validation across canopy-cover classes: high-canopy-cover plots exhibit lower prediction accuracy due to restricted USV visibility in Sentinel-2 imagery, whereas low-canopy-cover plots allow direct exposure of USV to the sensor, resulting in the lowest RMSE and highest variance explained across all models. This demonstrates that direct spectral visibility of USV provides the better prediction signal where overstorey obstruction is minimal. Previous literatures have combined Senitnel-2 images with satellite stereo image matching-based CHM and LiDAR CHM based texture variables for forest attribute mapping (Fassnacht et al. 2021). However, to the best of the authors’ knowledge, this is the first time such approaches have been applied to predict USV cover in boreal forests.

4.2 Value of structural information from CHM texture metrics

The contribution of CHM texture metrics to model performance was selective and strongly dependent on canopy cover class, with clear differences between training and validation behaviour. Across the training datasets, the addition of ALS- and Pléiades-derived CHM texture variables to Beta regression and RFR models produced only modest improvements, typically raising variance explained by 2–3% in Beta regression and up to 3% in RFR. These small but consistent gains indicate that structural metrics support spectral predictors but do not radically alter model fit under training conditions.

However, the validation results reveal a higher contrast, particularly for the RFR model incorporating Pléiades-derived CHM texture variables. This configuration achieved the best performance overall, with RMSE decreasing to 14.6 m² and variance explained increasing to 32% in low canopy cover plots far surpassing training performance (which explained only 9%). This pattern suggests that fine-scale structural heterogeneity, detectable in Pléiades CHM but not ALS CHM, introduces meaningful information that RFR can leverage to generalise well in open-canopy conditions.

In comparison, ALS-derived CHM texture variables produced moderate and more uniform improvements in both training and validation, typically increasing variance explained by 2–3% relative to spectral-only models. This reflects the high metric precision of ALS height data but also its coarser spatial grain, which limits its ability to describe horizontal texture relevant to understorey visibility. In high canopy cover classes, ALS CHM did not substantially enhance predictions, suggesting that structural height metrics alone cannot compensate for spectral occlusion by the overstorey.

The Beta regression coefficients further clarify how canopy structure influences the detectability of USV from spaceborne sensors. Negative coefficients for vr (canopy closeness) and CHM homogeneity indicate that uniform, closed canopies are associated with reduced spectral visibility of USV. Similarly, the negative effect of CHM dissimilarity suggests that fragmented or structurally cluttered canopy surfaces obscure the forest floor. In contrast, the positive coefficient for CHM contrast signals that greater vertical variability increases gap frequency, improving light penetration and enhancing USV detectability consistent with the previously identified spectral occlusion effect.

Taken together, the training and validation results show that CHM texture metrics contribute most when they capture canopy features that modulate understorey exposure, particularly in fine-resolution datasets and non-linear modelling frameworks.

4.3 Comparison between modelling approaches

Across both training and validation datasets, RFR consistently outperformed Beta regression in prediction accuracy, although the magnitude of improvement varied by canopy class and predictor set. In training, RFR produced slightly lower RMSE and marginally higher variance explained than Beta regression across most configurations. For example, Sentinel-2 spectral-only RFR models explained 1–7% of variance in training compared to 5–14% for Beta regression, but RFR generally provided more stable predictions across canopy classes.

However, the validation results showed a much stronger divergence, demonstrating the advantages of non-linear modelling for generalisation. RFR models incorporating Pléiades-derived CHM texture metrics achieved the highest accuracy, particularly in low canopy cover plots where explained variance increased to 32%, compared to only 1% for the equivalent Beta regression model. This indicates that the complex, non-linear interactions between canopy structure and spectral reflectance especially in fine-resolution CHM products are better captured by ensemble tree models.

Beta regression, while more interpretable, showed a clear tendency to degrade under the increased structural heterogeneity introduced by Pléiades CHM metrics. In training, its performance remained moderate (explaining up to 14% variance), but validation accuracy dropped for the Pléiades model, with RMSE rising to 23.1 m² and explained variance dropping to 1%. These results show that parametric approaches struggle with fine-grained structural complexity, leading to overfitting or insufficiently flexible model behaviour.

Overall, RFR demonstrated superior capacity to utilise CHM texture information, especially from high-resolution Pléiades data, but is prone to overfit training data. Beta regression, in contrast, provided consistent and interpretable coefficient behaviour that clarified ecological and radiometric drivers of USV cover, but at the cost of lower prediction performance.

5. Conclusions

This study demonstrates that USV cover can be predicted using only spaceborne spectral information with similar model performance as when canopy structural attributes derived from CHM texture metrics are incorporated. Both the training and validation results highlight the dominant role of canopy structure and occlusion to predict USV cover from spaceborne sensors. Training models typically explained 5–14% of variance (Beta regression) and 1–7% (RFR), reflecting the inherent difficulty of modelling USV beneath partially or fully closed canopies. Validation performance varied across canopy classes, with explained variance ranging from 0–32%, confirming that canopy openness is a primary factor influencing USV predictability using Sentinel-2.

Integrating CHM texture metrics improved performance, especially in non-linear modelling frameworks. The best-performing model, i.e., the RFR combined with Pléiades-derived CHM texture metrics achieved 32% explained variance in low canopy cover plots. This result indicates the importance of high-resolution structural detail. ALS CHM metrics provided more consistent contributions, but the model coefficients from Table 1 indicate that Pléiades-derived CHM texture metrics had higher impact on USV cover predictions.

Despite the overall modest prediction performance, the findings indicate that combining spectral and structural datasets yields meaningful gains in open or semi-open forest conditions. Future improvements in USV mapping may come from integrating very high spatial resolution remote sensing sources, including Pléiades satellite imagery, drone-derived photogrammetric point clouds, or high-density LiDAR, together with flexible non-linear learning algorithms that can fully exploit complex canopy–understorey interactions.

Acknowledgements

This study is part of the project “Prediction and Mapping of Understorey Vegetation Coverage and Change Analysis in Fennoscandian Ecosystem Using Satellite Remote Sensing Data”, funded by the Swedish National Space Agency (2024-00186). The authors would like to mention and thank Dr. Johannes Larson for acquisition of the field inventory data in the Krycklan catchment area.

Author contribution

Ritwika Mukhopadhyay: Writing – original draft, review, and editing, Visualization, Validation, Software, Methodology, Conceptualization, Investigation, Formal analysis, Data curation, Fund acquisition. **Ruben Valbuena:** Writing – review and editing. **Inka Bohlin:** Writing – review and editing, Methodology, Conceptualization.

Declaration of conflict of interest

The authors declare no conflict of interest that may have influenced the work.

References

Bohlin, Inka, Matti Maltamo, Henrik Hedenås, Tomas Lämås, Jonas Dahlgren, and Lauri Mehtätalo. 2021. ‘Predicting Bilberry and Cowberry Yields Using Airborne Laser Scanning and Other Auxiliary Data Combined with National Forest Inventory Field Plot Data’. *Forest Ecology and Management* 502:119737. doi:10.1016/J.FORECO.2021.119737.

Fassnacht, Fabian Ewald, Javiera Poblete-Olivares, Lucas Rivero, Javier Lopatin, Andrés Ceballos-Comisso, and Mauricio Galleguillos. 2021. ‘Using Sentinel-2 and Canopy Height Models to Derive a Landscape-Level Biomass Map Covering Multiple Vegetation Types’. *International Journal of Applied Earth Observation and Geoinformation* 94:102236. doi:10.1016/j.jag.2020.102236.

Larson, Johannes, Carl Vigren, Jörgen Wallerman, Anneli M. Ågren, Alex Appiah Mensah, and Hjalmar Laudon. 2024. ‘Tree Growth Potential and Its Relationship with Soil Moisture Conditions across a Heterogeneous Boreal Forest Landscape’. *Scientific Reports* 14(1):10611. doi:10.1038/s41598-024-61098-z.

Miina, Jari, Marius Hauglin, Aksel Granhus, Anne Linn Hykkerud, and Inger Martinussen. 2024. ‘Modelling and Mapping the Abundance of Lingonberry (*Vaccinium Vitis-Idaea* L.) in Norway’. *Global Ecology and Conservation* 54:e03195. doi:10.1016/j.gecco.2024.e03195.

Nilsson, Marie-Charlotte, and David A. Wardle. 2005. ‘Understory Vegetation as a Forest Ecosystem Driver: Evidence from the Northern Swedish Boreal Forest’. *Frontiers in Ecology* 3(8):421–28. doi:10.1890/1540-9295(2005)003.

Nilsson, Mats, Karin Nordkvist, Jonas Jonzén, Nils Lindgren, Peder Axensten, Jörgen Wallerman, Mikael Egberth, Svante Larsson, Liselott Nilsson, Johan Eriksson, and Håkan Olsson. 2017. ‘A Nationwide Forest Attribute Map of Sweden Predicted Using Airborne Laser Scanning Data and Field Data from the National Forest Inventory’. *Remote Sensing of Environment* 194:447–54. doi:10.1016/j.rse.2016.10.022.

Roberts, Dale, Norman Mueller, and Alexis Mcintyre. 2017. ‘High-Dimensional Pixel Composites From Earth Observation Time Series’. *IEEE Transactions on Geoscience and Remote Sensing* 55(11):6254–64. doi:10.1109/TGRS.2017.2723896.

Ståhl, G., S. Holm, Timothy G. Gregoire, T. Gobakken, E. Næsset, and R. Nelson. 2011. ‘Model-Assisted Estimation of Biomass in a LiDAR Sample Survey in Hedmark County, Norway’. *Canadian Journal of Forest Research* 41(1):83–95. doi:10.1139/X10-195.

Swedish National Forest Inventory, and Swedish Soil Inventory. 2021. *Riksinventeringen Av Skog*.

Appendix

Tables:

USV Beta regression	Cov	RMSE (m^2)	Variance explained
USV~S-2 spectral information (Eq. (4))	Overall High Med Low	20.6 20.4 23.0 17.3	9% 8% 5% 14%
USV~S-2 spectral information + ALS CHM texture variables (Eq. (5))	Overall High Med Low	20.6 19.9 22.7 18.4	11% 11% 8% 8%
USV~ S-2 spectral information + PL CHM texture variables (Eq. (5))	Overall High Med Low	21.0 20.4 23.3 18.2	9% 9% 5% 9%

Table 2. Statistical summary of training data (70% of the total dataset) for Eq. (4) and (5) fitted with Beta regression.

USV RFR	Cov	RMSE (m^2)	Variance explained
USV~S-2 spectral information (Eq. (4))	Overall High Med Low	22.7 21.9 26.1 18.3	1% 1% 3% 7%
USV~S-2 spectral information + ALS CHM texture variables (Eq. (5))	Overall High Med Low	22.3 22.1 24.5 18.6	1% 0.3% 0% 3%
USV~ S-2 spectral information + PL CHM texture variables (Eq. (5))	Overall High Med Low	21.9 21.5 24.9 17.3	3% 2% 0% 14%

Table 3. Statistical summary of training data (70% of the total dataset) for Eq. (4) and (5) fitted with random forest regression.

USV Beta regression	Cov	RMSE (m^2)	Variance explained
USV~S-2 spectral information (Eq. (4))	Overall High Med Low	21.7 20.5 24.9 16.3	5% 4% 1% 20%
USV~S-2 spectral information + ALS CHM texture variables (Eq. (5))	Overall High Med Low	22.2 20.9 24.8 18.2	4% 2% 2% 5%
USV~ S-2 spectral information + PL CHM texture variables (Eq. (5))	Overall High Med Low	23.3 20.4 25.9 23.1	3% 7% 1% 1%

Table 4. Statistical summary of validation data (30% of the total dataset) for Eq. (4) and (5) fitted with Beta regression.

USV RFR	Cov	RMSE (m^2)	Variance explained
USV~S-2 spectral information (Eq. (4))	Overall High Med Low	22.6 20.7 26.2 17.2	3% 5% 0% 10%
USV~S-2 spectral information + ALS CHM texture variables (Eq. (5))	Overall High Med Low	21.8 21.6 24.3 15.9	5% 0% 4% 21%
USV~ S-2 spectral information + PL CHM texture variables (Eq. (5))	Overall High Med Low	22.3 21.2 25.9 14.6	4% 2% 0% 32%

Table 5. Statistical summary of validation data (30% of the total dataset) for Eq. (4) and (5) fitted with random forest regression