https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Drone-based multispectral differentiation of subalpine vegetation at the treeline in the Southern Alps of New Zealand

Fabian Döweler[1*] and Martin Karl-Friedrich Bader[2*]

[1] Dragonfly Data Science, Wellington, New Zealand
[2] System Earth Science, Maastricht University, Brightlands Campus Greenport Venlo, The Netherlands

*Correspondence:
fab.doe@gmx.de, martin.bader@maastrichtuniversity.nl

Keywords
Vegetation index, unoccupied aerial vehicle (UAV), machine learning classifier, mountain biodiversity

Abstract
Subalpine ecosystems are highly dynamic environments that are particularly vulnerable to environmental change, yet their remote and rugged nature poses challenges for long-term monitoring. Unoccupied aerial vehicles (UAVs) equipped with multispectral sensors offer a scalable solution for high-resolution vegetation mapping in these landscapes. In this study, we integrated UAV-derived spectral data with machine learning (ML) classifiers to assess the effectiveness of different vegetation indices (VIs) in distinguishing subalpine plant communities. Principal component analysis (PCA) revealed that NDVI, SIPI2, MCARI, and CHL were highly correlated and strongly influenced the primary variance in the dataset, while NDRE and LCI contributed more evenly across principal components, and GNDVI was largely independent. Among the ML classifiers tested, extreme gradient boosting (XGBoost) achieved the greatest overall accuracy (81.3%) and Kappa (0.75), outperforming support vector machines (SVM) and random forest (RF). Classification confidence was highest for *Chionochloa* tussock (64.6–69.7%) and *Dracophyllum* scrub (70.6%), suggesting moderate reliability for these dominant vegetation types. Scrub and prostrate mat-forming communities exhibited lower classification accuracy, likely due to their heterogeneous canopy structure and greater spectral variability. The ecological boundaries of the subalpine zone, formed by *Fuscospora* forest and scree, were classified with high confidence, but the vegetation is dominated by tussock and shrubland. Feature importance analysis ranked NDVI, SIPI2, CHL, and MCARI highly in SVM and RF models, whereas LCI prevailed in XGBoost, underscoring how different algorithms leverage spectral information in classification tasks. These results emphasize the role of vegetation structure in classification accuracy, with dense, spectrally homogeneous vegetation types more reliably distinguished than mixed-species communities. Our study highlights UAV-based classification as a valuable tool for landscape-scale monitoring of subalpine vegetation. As UAV applications and ML workflows continue to evolve, optimizing classification approaches will enhance our ability to track ecological changes in subalpine and alpine regions worldwide.

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

1. Introduction

Globally, subalpine shrublands play a crucial role as biodiversity hotspots, supporting a wide range of endemic plant species and serving as critical habitats for various alpine-adapted fauna. Functioning as ecological transition zones, they facilitate interactions between species from lower and higher elevations, leading to unique assemblages. These ecosystems are also vital for carbon sequestration (Day et al., 2023), water regulation (Nicholls, 2023; Nicholls & Carey, 2021), and plant-soil nutrient balance (Urbina et al., 2020), exhibiting key ecosystem services in mountainous landscapes. Additionally, subalpine vegetation plays a fundamental role in global mountain ecosystems by acting as a crucial buffer against climate-driven changes (Hou, 2024) and microhabitat facilitation (Harsch et al., 2009). These environments are highly dynamic, shaped by the interplay of climate, topography, and ecological processes. However, ongoing climate change is reshaping subalpine ecosystems worldwide, altering species distributions, ecosystem functions, and landscape stability (Reid et al., 2022).

The vulnerability of alpine and subalpine vegetation to changing abiotic drivers is particularly concerning, as many species possess limited dispersal capacities and may be limited in their capability to respond to changes in their range limitations (Camac et al., 2021). Moreover, research on climate-driven treeline shifts (Körner, 2014) and microclimatic variability underscores the complexity and potential of subalpine ecosystems to modulate large scale abiotic drivers (Döweler et al., 2021, 2024) and biotic effects (e.g. control invasive species expansion; (Padalia et al., 2023). While some species may benefit from a warming climate, others, particularly alpine specialists, may not be able to compete with generalist species expanding their range from lower elevations (Thomas et al., 2023). This ecological reshuffling has profound implications for biodiversity, carbon storage, and ecosystem resilience, but often happens gradual and can only be thoroughly studied at the landscape scale, where large scale assessments of change of the subalpine lacks temporal and spatial resolution to adequately reflect these changes (Döweler et al., 2024).

In New Zealand, subalpine vegetation is characterized by a mosaic of tussock grasslands and low-stature shrubs, forming ecologically significant communities that influence ecosystem resilience and carbon storage (Mark, 2013; Day, 2023). The high cover of *Chionochloa* tussocks (Fig. 1), along with species such as *Dracophyllum uniflorum*, *Podocarpus nivalis*, and *Acrothamnus colensoi*, creates microclimates that buffer temperature extremes and support treeline regeneration of *Leuphozonia menziesii* and *Fuscospora cliffortioides* (Hook.f.) Oerst. (Döweler, 2021; Scherrer & Körner, 2010). These ecosystems are not only important for biodiversity but also provide a range of ecological functions, influencing water retention (van Galen et al., 2023) and soil carbon dynamics (Day et al., 2023).

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Figure 1 A typical subalpine belt vegetation composition in New Zealand (1365 m a above sea level, Craigieburn Valley, Arthurs Pass)

Despite their ecological significance, New Zealand's subalpine landscapes are experiencing complex transformations, yet our understanding of these changes remains limited. The lack of landscape-scale detailed vegetation mapping in the subalpine belt limits insights to capture these gradual but potentially significant shifts in over time (Day et al., 2023) in these often inaccessible regions. While treelines have shown limited upslope movement in response to warming (Harsch et al., 2012), subalpine vegetation composition is shifting due to factors such as woody encroachment and thickening, increased drought stress, and the potential expansion of invasive species (Chardon et al., 2024; Singh et al., 2024). Understanding how these communities are responding to climate change is crucial to further elucidate their role in stabilising these ecosystems to help inform conservation strategies (Reid, 2022; De Toma, 2025).

Remote sensing is a powerful tool for monitoring subalpine environments, where limited accessibility and the need for large-scale landscape assessments pose significant challenges (Walsh et al., 2009). It enables the classification of vegetation and detection of ecological shifts, offering a comprehensive perspective on mountain biodiversity and ecosystem dynamics. As climate change increasingly affects alpine and subalpine ecosystems, the ability to remotely assess vegetation composition across vast and often inaccessible areas has become a powerful method to study these ecosystems (Garbarino et al., 2023). Advances in high-resolution satellite and UAV-based remote sensing, combined with machine learning, have significantly improved vegetation classification, enhancing mapping accuracy and long-term monitoring (Mashiane et al., 2024; Nguyen et al., 2022). Access to light-weight sensors which can readily be mounted on increasingly affordable unoccupied aerial vehicles (UAV's) enables us to monitor subalpine ecotone in unprecedented detail using optical, multispectral, thermal, and LiDAR sensors (Döweler et al., 2024). Remote sensing technologies provide a robust means of tracking vegetation dynamics at ecologically meaningful scales, with satellite and aerial imagery proving effective in mapping subalpine vegetation and detecting temporal changes (De Toma et al., 2025). In some field studies, UAV-based deep learning methods may outperform human observers in delineating complex patterns in subalpine shrub communities (Moritake et al., 2024), endorsing their use for larger mapping endeavours of the subalpine in an approach to more accurately study vegetation shifts in response to climate change. These technological advancements offer critical insights for conservation planning and land management, ensuring more effective strategies for protecting subalpine ecosystems (Padalia et al., 2023).

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

This study investigates the potential of remotely sensed optical and multispectral vegetation indices to differentiate vegetation composition in a complex subalpine shrubland ecotone in New Zealand. Building on our previous classification and segmentation research in the Craigieburn Range (Arthur's Pass, New Zealand, Döweler et al., 2024), we aimed to compare the performance of three widely used machine learning classifiers (support vector machine, random forest and extreme gradient boosting) and, by extension, to identify which vegetation indices are most effective for distinguishing vegetation classes. This study aims to offer recommendations for vegetation indices and ML classifiers for future remote sensing applications in subalpine ecosystems. We hypothesised that classification accuracy will be highest in vegetation types with distinct spectral reflectance signatures and relatively low spectral and structural variability (e.g. sparsely vegetated scree, subalpine forest, tussock) and lowest in various types of scrub and prostrate mats where increased species interspersion may cause greater spectral overlap.

2 Material and Methods

2.1 Study site

The Craigieburn Valley study site (-43.111, 171.713) is located at 1365 metres above sea level on a southeast to southwest aspect within the eastern slopes of the Southern Alps, New Zealand (Fig. 2). The site is characterized by a montane to subalpine climate, with frequent frost events throughout the year (approximately 135 frost days annually) and an annual rainfall of around 1300 mm. The subalpine belt is dominated by *Chionochloa* tussock grasslands, interspersed with species such as *Dracophyllum uniflorum* Hook.f., *Podocarpus nivalis* Hook., and *Acrothamnus colensoi* (Hook.f.) Quinn, alongside areas of exposed scree. The adjacent treeline is formed by *Fuscospora cliffortioides* (Hook.f.) Oerst., marking the transition to the alpine treeline.

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Figure 2 Craigieburn study site (~ 4 ha). The subalpine belt has been covered by the drone survey. The visualisation shows an RGB coloured point cloud derived from a 2019 (May, 9 am austral winter) point cloud acquisition.

2.2 Subalpine vegetation classification and segmentation

For the classification and segmentation of the subalpine belt we we applied the same methodology as previously described in Döweler et al., 2024 producing a detailed map (0.05 m) of the subalpine vegetation, with low spatial offset (6.14 ± 4.03 cm; mean ± standard error) over an area of approximately 4 ha covering the supaline belt. We provide a brief summary of the method below, for the full workflow please see the respective paper.

For georeferencing, we used a differential GPS rover-base setup (Emlid Reach RS+) to locate ground control points (GCPs) consisting of chessboard-patterned panels or high-visibility spray markers, which were distributed across the UAV flight path for photogrammetric orthorectification. The differential GPS was also used to validate classification results by geolocating vegetation patches within the subalpine, resulting in over 600 ground-truth vegetation identifications in Craigieburn Valley. UAV imagery was captured using a DJI Phantom 4 (0.01 m RGB resolution) and a Parrot Sequoia+ multispectral sensor (0.05 m resolution), mounted for aerial surveys during the 2018/19 austral growing season. Pre- and post-flight calibration followed the One-Point Calibration plus Sunshine Sensor method using the Parrot target plate. The UAV flight paths were planned using UgCS software (SPH Engineering, 2025), incorporating an 8 m resolution digital terrain model

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

(Geographx, 2016) for altitude control, with a 5 m/s flight speed and 80% along-track overlap for photogrammetric processing.

Post-processing was performed in Pix4D 9.4 (Pix4D, 2025) for image orthorectification and spatial alignment using mapped GCPs. RGB and multispectral datasets were aligned manually and projected into the New Zealand Transverse Mercator (NZTM 2000) coordinate system. Object-based image analysis was conducted using eCognition Developer 9.0 (Trimble, 2025), applying multiresolution segmentation (scale: 250, shape: 0.1, compactness: 0.7) to generate spectrally homogeneous objects. A nearest neighbor classifier, trained on geolocated field data, was used to categorize land cover into five vegetation types and scree. Feature selection was optimized through the Feature Space Optimization tool in eCognition, maximizing vegetation separability across spectral bands. Classification accuracy was evaluated using an error confusion matrix based on the 600 ground-truth points, with Kappa statistics indicating an overall accuracy of 89.7% (Fig. 3, Döweler et al., 2024)

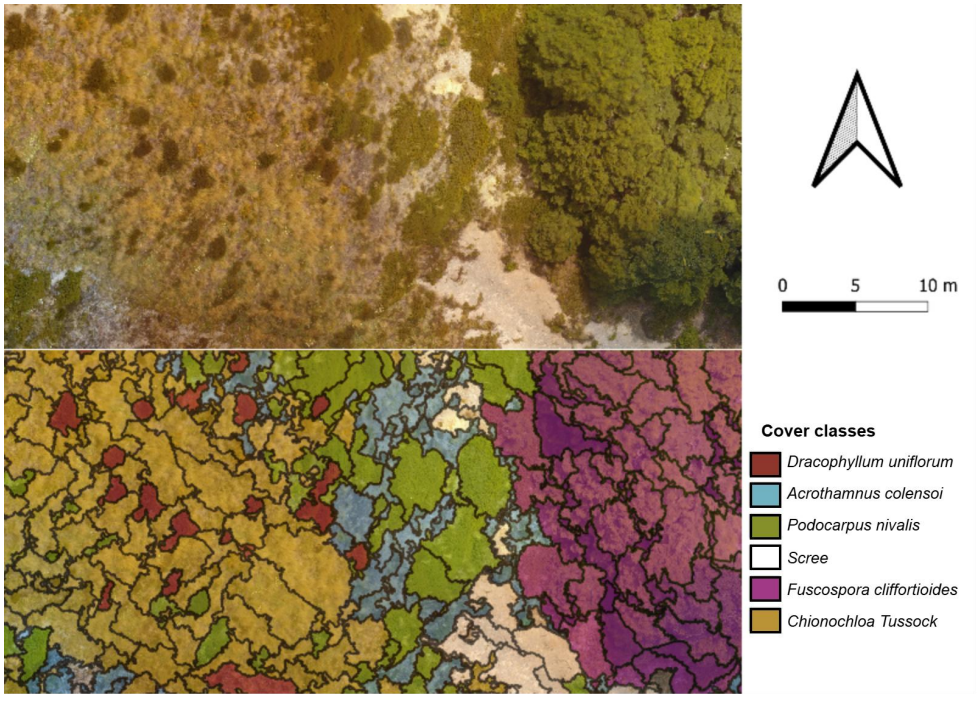

Figure 3 Results for the pre-existing subalpine vegetation segmentation and classification for the Craigieburn valley established by Döweler et al. (2024)

2.3 Vegetation indices used to discern cover classes

Following the segmentation and classification of the vegetation cover classes, we derived a suite of vegetation indices from the available multispectral bands to test their capability in discerning cover classes. These indices capture plant functional traits that influence

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

productivity, stress responses, and spectral variability across different vegetation types. By leveraging multispectral reflectance data, we aimed to improve vegetation classification accuracy and investigate how vegetation properties vary across topographic gradients.

All UAV-derived spectral data was sampled to a uniform 0.01 m resolution using a bicubic interpolation in GDAL (3.4.1; Rouault et al., 2025). Vegetation classification polygons were processed in GeoPandas (1.0.1; Van den Bossche et al., 2024) by extracting their centroids, which were then buffered by 0.2 m to create localized sampling zones. This buffering step minimized the influence of mixed-pixel effects and ensured that spectral values extracted within each zone were representative of potential spectral heterogeneity within a single vegetation type. We used a spatial filtering GeoPandas, ensuring that each buffered centroid remained entirely within its original vegetation class. After filtering, the final vegetation sample counts were 4436 *Fuscospora cliffortioides*, 1809 scree, 1384 *Chionochloa* tussock (-12%), 883 *Acrothamnus colensoi* (-13%), 676 *Podocarpus nivalis* (-10%), and 293 *Dracophyllum* spp. (-3%), with a total of 9476 vegetation samples retained for analysis. The filtered dataset was used to compute zonal statistics for each vegetation index using the rasterstats (0.15.0; Perry, 2025) package, extracting median, mean and standard deviation values within each segmented vegetation type.

The vegetation indices used in this analysis (Table 1) are widely applied in remote sensing analyses on vegetation health and growth. We included the Normalized Difference Vegetation Index (NDVI), Structure Insensitive Pigment Index 2 (SIPI2), Modified Chlorophyll Absorption Ratio Index (MCARI), Green Normalized Difference Vegetation Index (GNDVI), Chlorophyll Index (CHL), Normalized Difference Red Edge Index (NDRE), and Leaf Chlorophyll Index (LCI). NDVI was calculated to assess overall vegetation health, while SIPI2 was included as an indicator of vegetation stress and pigment ratios. MCARI, GNDVI, CHL, and LCI were shifting the focus on chlorophyll concentrations and photosynthetic potential, while the NDRE has been selected to provide a focus on red-edge detectable indication of early stress.

Table 1 Overview of vegetation indices and respective band calculations used in the current study, which represent unpublished data from a previous remote sensing study in the same area (Döweler et al., 2024). Green (550 nm), Red (660 nm), Near-infrared (790 nm, NIR), Red Edge (735 nm, RE) derived from the Parrot Sequoia, Blue band (450 nm) extracted from the Phantom 4 RGB sensor.

Index	Name	Calculation	Reference
NDVI	Normalized Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$	Rouse et al., 1973
SIPI2	Structure Insensitive Pigment Index 2	$\frac{NIR - Blue}{NIR - Red}$	Peñuelas et al., 1995
MCARI	Modified Chlorophyll Absorption Ratio Index	$[(RE - Red) - 0.2 \times (Red\ Edge - Green)] \times \frac{RE}{Red}$	Daughtry et al., 2000

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Index	Name	Calculation	Reference
GNDVI	Green Normalized Difference Vegetation Index	$\frac{NIR - Green}{NIR + Green}$	Gitelson & Merzlyak, 1998
CHL	Chlorophyll Red-Edge Index	$\frac{NIR}{RE} - 1$	Gitelson et al., 2003
NDRE	Normalized Difference Red Edge Index	$\frac{NIR - RE}{NIR + RE}$	Barnes et al., 2000
LCI	Leaf Chlorophyll Index	$\frac{NIR - RE}{NIR + Red}$	Haboudane et al., 2002

2.4 Statistical analysis

All statistical computations and graphics were performed using the R software within the RStudio integrated development environment (R version 4.4.1, R Core Team, 2024, RStudio version 2024.09.0+375, Posit team, 2024). To provide a general overview, we performed a principal component analysis (PCA) based on the overall medians of the vegetation indices (i.e. the median of the sample medians of each index). The PCA results were visualised in a biplot to illustrate similarities among vegetation types and the contribution of VIs to the principal components (loading vectors) as well as their interrelationships.

We used three popular machine learning approaches to classify the six vegetation types based on the centroid buffered medians of seven features (vegetation indices): NDVI, GNDVI, CHL, LCI, MCARI, NDRE, SIPI2. We created a balanced 80/20 training-to-test split of our data, ensuring random sampling within each class to preserve the overall class distribution (balanced splits are obtained by providing a factor, i.e. the vegetation type labels, to the *createDataPartition* function, R package *caret*, Kuhn, 2008)
We trained a Support Vector Machine (SVM) classifier using a radial basis function (RBF) kernel to differentiate between vegetation types (R package *e1071*, Meyer et al., 2024)
To address class imbalances, we assigned class weights inversely proportional to their frequencies (following the approach in the Python package *scikit-learn*, Pedregosa et al., 2011). Hyperparameter optimization was performed via 5-fold cross-validation, selecting the optimal values of C (regularization parameter) and γ (kernel coefficient, controlling the influence of data points) using a grid search. The final model featured an RBF kernel with C = 10 and γ = 0.01.
We also ran a random forest approach to classify the vegetation types (R package *randomForest*, Liaw & Wiener, 2002). The model algorithm was run with the default 500 number of trees and a hyperparameter tuning procedure suggested two randomly sampled features (predictors) at each split (mtry = 2). Stratified sampling was used to ensure that each tree was trained on a random sample containing observations from all vegetation types.

In addition, we applied extreme gradient boosting (function *xgb.train* in R package *xgboost*, Chen et al., 2024) (Chen et al., 2024) for vegetation type classification. A grid search was used for hyperparameter tuning (final hyperparameter settings: eta = 0.05, max_depth = 6, gamma = 2) and 5-fold cross-validation to determine the optimal number of iterations (100 cross-validation runs allowing a maximum of 500 iterations yielded a mean of 150 iterations).

For all three machine learning approaches, the model performance scores were derived from a confusion matrix contrasting true and predicted class labels (*confusionMatrix* function

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

in R package *caret*, Kuhn, 2008). Class-level sensitivity (recall) and specificity scores were each averaged to an overall score. In addition, the multi-class area under the receiver operating characteristic curve (AUC_{mc}) was calculated as the mean of the class-specific AUCs (function *multiclass.roc* in R package *pROC*, Robin et al., 2011). Finally, we applied a permutation-based feature importance analysis to all three classifiers. This feature importance procedure was run with 30 permutation rounds using a cross-entropy loss function to evaluate feature contribution (*feature_importance* function in R package *ingredients* relying on the cross-entropy loss function in R package *DALEX*, Biecek, 2018; Biecek et al., 2023).

3 Results

3.1 Principal component analysis (PCA)

The PCA showed that PC1 explained roughly 95 % and PC2 nearly 5 % of the variation in the aggregated data. In the PCA biplot, *Podocarpus* scrub and *Dracophyllum* scrub formed the only discernible cluster (which could perhaps include *Chionochloa* tussock), indicating similar profiles related to the VIs. By contrast, the remaining vegetation types showed distinct profiles (Fig. 4). The small angular distances between the loading vectors of NDVI, SIPI2, MCARI and CHL suggest that these indices were all positively correlated and their fairly horizontal alignment indicates a strong influence on PC1 (Fig. 4). The LCI and NDRE were also positively correlated and contributed roughly equally to both PCs. The GNDVI had a strong negative contribution to PC2, and was weakly or uncorrelated with the other VIs considering the large angular distances to the other loading vectors (Fig. 4).

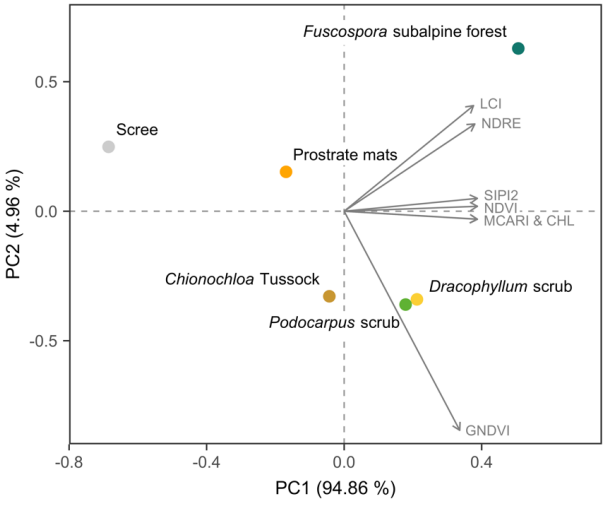

Figure 4 Principal component analysis (PCA) biplot displaying the scores associated with various subalpine vegetation types (filled circles) and the contribution of vegetation indices (arrows indicating loading vectors) to the principal components. The vegetation indices were scaled prior to PCA. Please note that the eigenvectors of the MCARI and CHL indices are virtually equal.

3.2 Machine learning classifiers

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Based on common performance metrics (accuracy, Cohen's Kappa statistic, AUC, sensitivity and specificity), the support vector machine (SVM) and random forest (RF) classifiers performed similarly well but were slightly outperformed by the extreme gradient boosting algorithm (XGBoost) (Fig. 5). Accuracy was just below 80 % in the SVM and the RF, while reaching 81.3 % in the XGBoost approach. Given our unbalanced data, the more robust Kappa statistic seems more informative than overall accuracy. Cohen's Kappa was highest in the XGBoost classifier at 0.75, compared to 0.73 in the SVM and 0.72 in the RF approach (Fig. 5). The multi-class AUC-ROC and specificity were identical in all three algorithms (AUC_{mc} = 0.93, specificity = 0.96), while sensitivity (recall) varied from 0.72 in the SVM and 0.70 in the XGBoost to 0.68 in the RF classifier (Fig. 5). As judged by the percentage of correct classifications in the confusion matrices of the three classifiers, *Fuscospora* subalpine forest (93.3 – 94.7 %) and scree (86.7 – 88.7 %) can be identified with high to very high confidence, followed by moderate classification confidence for *Chionochloa* Tussock (64.6 – 69.7 %). Classification confidence for the remaining vegetation types was mostly low (< 60 %), except for the SVM's 70.6 % correct classifications of *Dracophyllum* scrub and its 63.2 % accuracy in classifying *Acrothamnus colensoi* (Fig. 5).

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

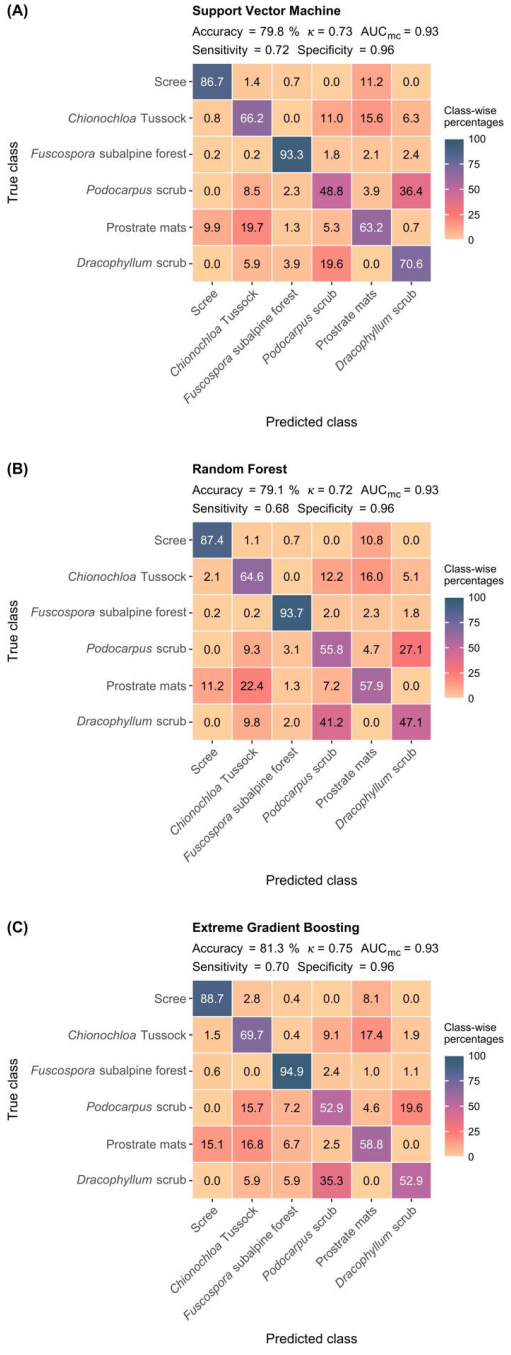

Figure 5 Heatmaps of the confusion matrices of a support vector machine (**A**) a random forest model (**B**) and an extreme gradient boosting machine (**C**). κ indicates Cohen's Kappa statistic, AUC_{mc} denotes the multi-class area under the ROC curve. Accuracy, κ and AUC_{mc} are overall model statistics, while sensitivity (recall) and specificity indicate averages of the class-specific metrics.

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

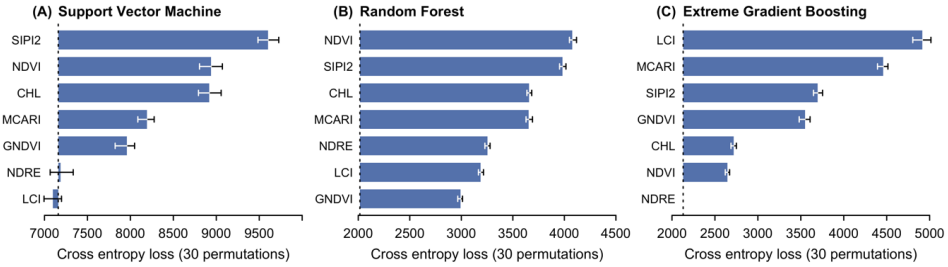

Figure 6 Permutation-based feature importance (variable importance) of a support vector machine (**A**), a random forest model (**B**), and an extreme gradient boosting model (**C**). The vertical dotted lines indicate the cross-entropy loss of the full model (no variables dropped). The error bars signify the 2.5th and 97.5th percentiles of the 30 permutations.

In the SVM algorithm, SIPI2 represented the most important variable, followed by the equally relevant NDVI and CHL, while MCARI and GNDVI were of moderate importance (Fig. 6 A). NDRE and LCI had little to no significance in the SVM. In the RF model, NDVI and SIPI2 were similarly important, followed by equal contributions from CHL and MCARI, and to a lesser but similar extent from NDRE and LCI, with GNDVI coming in with a slightly lower entropy loss value (Fig. 6 B). By contrast, the XGBoost algorithm relied most heavily on LCI, followed by MCARI and, on a similar level, SIPI2 and GNDIV (Fig. 6 C). CHL and NDVI were of lesser importance, while NDRE represented an entropy-irrelevant feature.

4. Discussion

Mountain ecosystems support a high level of plant diversity and endemism, especially at high elevation but coordinated monitoring efforts are scarce (Perrigo et al., 2020), often hampered by limited access. In the future, more affordable drone technology will bolster remote sensing based mapping and monitoring of these hard-to-reach ecosystems, and information that aids in optimising classification accuracy will facilitate progress in this field. Here, we compared the performance of three popular ML classifiers and assessed their feature importance to rank common VIs based on their impact on the classification algorithm.

Judged by the overall accuracy and Cohen's Kappa, the XGBoost algorithm differentiated the existing vegetation types best, but SVM and RF both performed nearly as well. Consistent with our findings, XGBoost outperformed SVM and RF in urban land use-land cover and forest classification tasks (Georganos et al., 2018; Ramdani & Furqon, 2022). In line with our hypothesis, spectrally more distinct vegetation types with lower spectral and structural variation showed the greatest proportion of correct classifications in the confusion matrices of the three classifiers, which is consistent with previous findings reported for diverse wetland ecosystems (Schmidt & Skidmore, 2003). In our study, these spectrally distinct vegetation types included *Fuscospora cliffortioides* subalpine forest and *Chionochloa* tussock, which are characterised by dense, uniform foliage and/or well-defined canopy structures, making them easier to differentiate in the ML classification process (Ollinger, 2011). The sparsely vegetated scree is characterised by a relatively uniform spectral signature, i.e. spectrally bland, which facilitates classification.

The importance of the used VIs was ranked similarly in the SVM and RF but differed greatly in the XGBoost algorithm, suggesting that the different ML approaches rely on distinct spectral properties for classification. The SVM and the RF model agreed closely in regard to

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

the four most important VIs (SIPI2, NDVI, CHL, MCARI). These VIs reflect pigment content and general vegetation vigor and are closely related to photosynthetic activity, which makes them suitable for distinguishing between broad vegetation types. By contrast, in the XGBoost approach LCI emerged as the most important feature, which had minimal leverage in the other two classifiers. This discrepancy likely reflects XGBoost's greedy search for splits that minimise the loss function, allowing a feature to gain importance even if it is of little relevance in SVM and RF algorithms (Kamdem and Fokoue, 2022). LCI's top ranking in XGBoost suggests it may hold critical information for distinguishing subtle variations in leaf structure or pigment content that were not prioritized in the RF and SVM algorithms.
A recent review on plant and vegetation classification based on spectral signatures revealed that besides the biological properties also the methodological approaches, the scale at which the recordings are performed and not least the applied feature selection procedure itself may all have a strong influence on feature importance in ML classifiers (Hennessy et al., 2020). To eleminate the latter source of variation, we applied the same permutation-based feature importance analysis based on cross-entropy loss to all three classifiers (Biecek, 2018; Biecek and Baniecki, 2023).

Another notable finding is the weak contribution of NDRE in all classifiers, indicating its limited role in distinguishing subalpine vegetation types for the subalpine ecotone at our study site, which calls for verification in other high-elevation transition zones. Unlike NDVI and MCARI, which are widely used in vegetation classification, NDRE is often associated with deeper canopy penetration and is particularly useful in detecting nitrogen stress and subtle variations in chlorophyll content (Boiarskii & Hasegawa, 2019). The lack of importance of NDRE in this study suggests that these characteristics were not primary drivers of spectral separability among the subalpine vegetation types analyzed here. Instead, indices related to general canopy structure and pigment concentration (such as NDVI, CHL, and MCARI) proved more effective.

The results of our study highlight the critical role of vegetation structure in classification accuracy, with dense, spectrally uniform vegetation types being more reliably identified than structurally diverse shrublands and mixed-species communities. As UAV technology becomes increasingly accessible, further refinement of vegetation index selection and classification methodologies is essential to capture the often subtle responses of subalpine vegetation to abiotic stressors, which are being exacerbated by climate change. The decline of certain species could lead to the loss of critical microhabitats and climatic niches, which serve as stepping stones for the recruitment of subalpine specialists and treeline forming species (Döweler et al., 2021; Frei et al., 2018; Harsch et al., 2009). For New Zealand, a landscape-scale classification of the subalpine can support monitoring the impact of invasive herbivores on these ecosystems, as their grazing pressure threatens both vegetation dynamics and the region's carbon sequestration potential (Lee et al., 2000). Advancements in remote sensing and machine learning offer novel pathways to improve monitoring efforts, enabling us to more clearly formulate and track conservation targets.

5. Conclusion

The effective integration of vegetation indices with modern ML classifiers presents a powerful tool for tracking ecological shifts, particularly in remote and rugged environments. As UAV operations become more affordable, their application in long-term monitoring will be invaluable for detecting and understanding vegetation changes in otherwise inaccessible regions. Expanding these efforts through global collaboration will provide deeper insights into the poorly understood dynamics of subalpine ecosystems under changing climatic conditions. Given the crucial role of subalpine grasslands and woody vegetation in carbon sequestration (Ward et al., 2014) and other ecosystem services such as maintaining biodiversity, erosion protection, runoff regulation and snow retention, their ecological

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

trajectories must be closely monitored to inform conservation strategies aimed at mitigating species loss and preserving ecosystem functions.

Data availability

All data for this publication can be requested from the corresponding author.

Author contributions

FD contributed to data acquisition and sample collection and data preparation. MB contributed to data analysis. Both authors contributed equally to the writing of this manuscript.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

Support for this research was provided by Auckland University of Technology (AUT, Auckland). We would like to thank the Department of Conservation (DOC) for access permissions, as well as Lincoln University (Canterbury, New Zealand) for providing accommodation.

References

Barnes, E., Clarke, T., Richards, S., Colaizzi, P., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., & others. (2000). *Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data*. *1619*, 6.

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, *19*(84), 1–5.

Biecek, P., Baniecki, H., & Izdebski, A. (2023). *ingredients: Effects and Importances of Model Ingredients* (Version 2.3.0) [Computer software]. https://cran.r-project.org/web/packages/ingredients/index.html

Boiarskii, B., & Hasegawa, H. (2019). Comparison of NDVI and NDRE indices to detect differences in vegetation and chlorophyll content. *J. Mech. Contin. Math. Sci*, *4*, 20–29.

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

482 Camac, J. S., Umbers, K. D. L., Morgan, J. W., Geange, S. R., Hanea, A., Slatyer, R. A.,
483 McDougall, K. L., Venn, S. E., Vesk, P. A., Hoffmann, A. A., & Nicotra, A. B. (2021).
484 Predicting species and community responses to global change using structured
485 expert judgement: An Australian mountain ecosystems case study. *Global Change*
486 *Biology*, *27*(18), 4420–4434. https://doi.org/10.1111/gcb.15750
487 Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R.,
488 Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., &
489 implementation), Xgb. contributors (base Xgb. (2024). *xgboost: Extreme Gradient*
490 *Boosting* (Version 1.7.8.1) [Computer software]. https://cran.r-
491 project.org/web/packages/xgboost/index.html
492 Daughtry, C. S. T., Walthall, C. L., Kim, M. S., de Colstoun, E. B., & McMurtrey, J. E. (2000).
493 Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance.
494 *Remote Sensing of Environment*, *74*(2), 229–239. https://doi.org/10.1016/S0034-
495 4257(00)00113-9
496 Day, N., Lavorel, S., Barratt, B., Dickinson, K., Buckley, H., Christensen, B., Curran, T., &
497 Norton, D. (2023). Predicting ecological change in tussock grasslands of Aotearoa
498 New Zealand. *New Zealand Journal of Ecology*.
499 https://doi.org/10.20417/nzjecol.47.3549
500 De Toma, A., Malavasi, M., Marzialetti, F., & Cutini, M. (2025). Unveiling spatial patterns and
501 trajectories of shrub dynamics in Mediterranean alpine ecosystems. *Plant Ecology*,
502 *226*(2), 149–160. https://doi.org/10.1007/s11258-024-01479-6
503 Döweler, F. (2021). *Causes of Recruitment Limitation at Abrupt Alpine Treelines* [Auckland
504 University of Technology]. https://hdl.handle.net/10292/14038
505 Döweler, F., Case, B. S., Buckley, H. L., & Bader, M. K.-F. (2021). High light-induced
506 photoinhibition is not limiting seedling establishment at abrupt treeline ecotones in
507 New Zealand. *Tree Physiology*, *41*(11), 2034–2045.
508 https://doi.org/10.1093/treephys/tpab061
509 Döweler, F., Fransson, J. E. S., & Bader, M. K.-F. (2024). Linking High-Resolution UAV-

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Based Remote Sensing Data to Long-Term Vegetation Sampling—A Novel Workflow to Study Slow Ecotone Dynamics. *Remote Sensing*, *16*(5), Article 5. https://doi.org/10.3390/rs16050840

Frei, E. R., Bianchi, E., Bernareggi, G., Bebi, P., Dawes, M. A., Brown, C. D., Trant, A. J., Mamet, S. D., & Rixen, C. (2018). Biotic and abiotic drivers of tree seedling recruitment across an alpine treeline ecotone. *Scientific Reports*, *8*(1), 10894. https://doi.org/10.1038/s41598-018-28808-w

Garbarino, M., Morresi, D., Anselmetto, N., & Weisberg, P. J. (2023). Treeline remote sensing: From tracking treeline shifts to multi-dimensional monitoring of ecotonal change. *Remote Sensing in Ecology and Conservation*, *n/a*(n/a). https://doi.org/10.1002/rse2.351

Geographx. (2016). *LINZ NZ 8m Digital Elevation Model 2016*. https://data.linz.govt.nz/layer/51768-nz-8m-digital-elevation-model-2012/

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geoscience and Remote Sensing Letters*, *15*(4), 607–611. IEEE Geoscience and Remote Sensing Letters. https://doi.org/10.1109/LGRS.2018.2803259

Gitelson, A. A., Gritz †, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*(3), 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A. A., & Merzlyak, M. N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, *22*(5), 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*(2), 416–

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

426. https://doi.org/10.1016/S0034-4257(02)00018-4

Harsch, M. A., Hulme, P. E., McGlone, M. S., & Duncan, R. P. (2009). Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology Letters*, *12*(10), 1040–1049. https://doi.org/10.1111/j.1461-0248.2009.01355.x

Körner, C. (2014). *Mountain ecosystems in a changing environment*. ResearchGate. http://dx.doi.org/10.1553/ecomont-6-1s71

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Lee, W. g., Fenner, M., Loughnan, A., & Lloyd, K. m. (2000). Long-term effects of defoliation: Incomplete recovery of a New Zealand alpine tussock grass, Chionochloa pallens, after 20 years. *Journal of Applied Ecology*, *37*(2), 348–355. https://doi.org/10.1046/j.1365-2664.2000.00498.x

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. *2*.

Mashiane, K., Ramoelo, A., & Adelabu, S. (2024). Prediction of species richness and diversity in sub-alpine grasslands using satellite remote sensing and random forest machine-learning algorithm. *Applied Vegetation Science*, *27*(2), e12778. https://doi.org/10.1111/avsc.12778

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chih-Chung, C. (libsvm, & Chih-Chen, L. (2024). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien* (Version 1.7-16) [Computer software]. https://cran.r-project.org/web/packages/e1071/index.html

Moritake, K., Cabezas, M., Nhung, T. T. C., Lopez Caceres, M. L., & Diez, Y. (2024). Sub-alpine shrub classification using UAV images: Performance of human observers vs DL classifiers. *Ecological Informatics*, *80*(102462), 1–16. https://doi.org/10.1016/j.ecoinf.2024.102462

Nguyen, T.-A., Kellenberger, B., & Tuia, D. (2022). Mapping forest in the Swiss Alps treeline ecotone with explainable deep learning. *Remote Sensing of Environment*, *281*, 113217. https://doi.org/10.1016/j.rse.2022.113217

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

Nicholls, E. (2023). *Resolving the role of subarctic vegetation on mountain water cycling in a rapidly changing climate* [Thesis]. https://macsphere.mcmaster.ca/handle/11375/28472

Nicholls, E., & Carey, S. K. (2021). Evapotranspiration and energy partitioning across a forest-shrub vegetation gradient in a subarctic, alpine catchment. *Journal of Hydrology*, *602*, 126790. https://doi.org/10.1016/j.jhydrol.2021.126790

Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, *189*(2), 375–394.

Padalia, H., Rai, I. D., Pangtey, D., Rana, K., Khuroo, A. A., Nandy, S., Singh, G., Sekar, K. C., Sharma, N., Uniyal, S. K., Talukdar, G., Saran, S., Chandra, N., Bushra, A., Ahmad, R., Thakar, A., Deepak, Stanzin, J., & Rawat, G. S. (2023). Fine-scale classification and mapping of subalpine-alpine vegetation and their environmental correlates in the Himalayan global biodiversity hotspot. *Biodiversity and Conservation*, *32*(13), 4387–4423. https://doi.org/10.1007/s10531-023-02702-y

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, *12*(null), 2825–2830.

Peñuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, *31*(2), 221–230.

Perrigo, A., Hoorn, C., & Antonelli, A. (2020). Why mountains matter for biodiversity. *Journal of Biogeography*, *47*(2), 315–325. https://doi.org/10.1111/jbi.13731

Perry, M. (2025). *Perrygeo/python-rasterstats* [Python]. https://github.com/perrygeo/python-rasterstats (Original work published 2013)

Pix4D. (2025). *Professional photogrammetry and drone mapping software*. Pix4D. https://www.pix4d.com/

Ramdani, F., & Furqon, M. T. (2022). The simplicity of XGBoost algorithm versus the

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

complexity of Random Forest, Support Vector Machine, and Neural Networks algorithms in urban forest classification. *F1000Research*, *11*, 1069.

Reid, K. A., Reid, D. G., & Brown, C. D. (2022). Patterns of vegetation change in Yukon: Recent findings and future research in dynamic subarctic ecosystems. *Environmental Reviews*, *30*(3), 380–401. https://doi.org/10.1139/er-2021-0110

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77. https://doi.org/10.1186/1471-2105-12-77

Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., Szekeres, T., Tourigny, E., Landa, M., Miara, I., Elliston, B., Chaitanya, K., Plesea, L., Morissette, D., Jolma, A., Dawson, N., Baston, D., de Stigter, C., & Miura, H. (2025). *GDAL* (Version v3.10.2) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.5884351

Rouse, J. W., Haas, R. H., Deering, D. W., & Schell, J. A. (1973). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (RSC-1978-2). https://ntrs.nasa.gov/citations/19740004927

Scherrer, D., & Körner, C. (2010). Infra-red thermometry of alpine landscapes challenges climatic warming projections. *Global Change Biology*, *16*(9), 2602–2613. https://doi.org/10.1111/j.1365-2486.2009.02122.x

Schmidt, K., & Skidmore, A. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, *85*(1), 92–108.

SPH Engineering. (2025). *UgCS - Drone flight planning software*. https://www.sphengineering.com/flight-planning/ugcs

Thomas, A., Meudt, H. M., Larcombe, M. J., Igea, J., Lee, W. G., Antonelli, A., & Tanentzap, A. J. (2023). Multiple origins of mountain biodiversity in New Zealand's largest plant radiation. *Journal of Biogeography*, *50*(5), 947–960. https://doi.org/10.1111/jbi.14589

Trimble. (2025). *eCognition*. https://geospatial.trimble.com/products/software/trimble-ecognition

https://doi.org/10.5194/egusphere-2025-926
Preprint. Discussion started: 10 March 2025

622 Urbina, I., Grau, O., Sardans, J., Ninot, J. M., & Peñuelas, J. (2020). Encroachment of
623 shrubs into subalpine grasslands in the Pyrenees changes the plant-soil
624 stoichiometry spectrum. *Plant and Soil*, *448*(1), 37–53.
625 https://doi.org/10.1007/s11104-019-04420-3
626 Van den Bossche, J., Jordahl, K., Fleischmann, M., Richards, M., McBride, J., Wasserman,
627 J., Garcia Badaracco, A., Snow, A. D., Ward, B., Tratner, J., Gerard, J., Perry, M.,
628 cjqf, Hjelle, G. A., Taves, M., ter Hoeven, E., Cochran, M., Bell, R., rraymondgh, …
629 Gardiner, J. (2024). geopandas/geopandas: V1.0.1. *Zenodo*.
630 https://doi.org/10.5281/zenodo.12625316
631 van Galen, L. G., Lord, J. M., Orlovich, D. A., & Larcombe, M. J. (2023). The effect of
632 stratification and shade on germination of Nothofagus cliffortioides (Hook.f.) Oerst
633 seeds. *New Zealand Journal of Botany*, *61*(4), 351–360.
634 https://doi.org/10.1080/0028825X.2022.2156355
635 Walsh, S. J., Brown, D. G., Geddes, C. A., Weiss, D. J., McKnight, S., Hammer, E. S., &
636 Tuttle, J. P. (2009). Chapter 2 Pattern–Process Relations in the Alpine and Subalpine
637 Environments: A Remote Sensing and GIScience Perspective. In D. R. Butler, G. P.
638 Malanson, S. J. Walsh, & D. B. Fagre (Eds.), *Developments in Earth Surface*
639 *Processes* (Vol. 12, pp. 11–34). Elsevier. https://doi.org/10.1016/S0928-
640 2025(08)00202-2
641 Ward, A., Dargusch, P., Thomas, S., Liu, Y., & Fulton, E. A. (2014). A global estimate of
642 carbon stored in the world's mountain grasslands and shrublands, and the
643 implications for climate policy. *Global Environmental Change*, *28*, 14–24.
644 https://doi.org/10.1016/j.gloenvcha.2014.05.008