Responses to Reviewer #1

Main Comment

Reviewer: The authors state that they used "two cloud-free WorldView-2 orthorectified images with a spatial resolution of 0.4 meters, acquired on November 3, 2012, and September 26, 2021." However, they later clarify that only the panchromatic (PAN) band is available at this resolution, while they appear to use the color bands instead. This is unclear—did they use pansharpening? Please clarify which bands were actually used, at what resolution, and whether pansharpening was applied.

Response: We appreciate this critical observation. Our study obtained the 8-band multispectral WorldView-2 imagery, which originally had a spatial resolution of 1.64 m, along with the 0.41 m panchromatic image. We applied a pansharpening process to the multispectral bands to enhance spatial detail, resulting in an effective spatial resolution of 0.4 m for all used spectral bands. This pansharpened dataset was used for all feature extraction, including vegetation indices and texture features. We revised the manuscript as follows:

> "To enhance spatial detail, all multispectral bands were pansharpened using the corresponding high-resolution panchromatic band, yielding a uniform spatial resolution of 0.4 meters across all datasets used for feature extraction. The pansharpened multispectral imagery was the basis for deriving vegetation indices and texture features."

Reviewer: The origin of the training data is not clearly explained. The authors write: "Ground truth data in the study area were labeled using a pixel-based approach and categorized into four classes: bare land, forest, krummholz, and shadow (Fig. 3)." Does this mean an operator manually classified these images? If both images were already classified, what is the purpose of the complex processing workflow? Were both images used for training? If only one image was used for training, why would we expect the same classification accuracy to transfer to the second image, especially given possible environmental and seasonal differences?

Response: Yes, the ground truth was manually labeled and validated by trained operators and domain experts using a pixel-based approach (Fig. 3) in the 2021 WorldView-2 orthorectified images. The 2021 images were used for model training and validation, while the 2012 images were used for model evaluation. The classification results of 2012 and 2021 images were compared for temporal change analysis.

The purpose of the complex processing workflow was hoping to establish a universal model that can provide reliable alpine treeline ecotone classification. Only 2021 images were used for model training. We did not expect to see the same classification accuracy. Instead, we evaluated the 2012 image classification results carefully and the results were used to study the temporal and spatial changes.

The environmental and seasonal differences were minimized by using the autumn images when vegetation in the area had entered dormancy with less phenological presentation such as flowering or leaf flushing.

Additionally, we applied histogram matching during preprocessing to reduce radiometric and color inconsistencies caused by differences in lighting and atmospheric conditions. We revised the manuscript as follows:

"Two orthorectified, cloud-free WorldView-2 images acquired on November 3, 2012, and September 26, 2021, were obtained from RiChi Technology Co., Ltd. (New Taipei City, Taiwan). Both images were captured in the autumn season when vegetation had entered dormancy, minimizing the influence of phenological variability such as flowering. Histogram matching was applied to ensure radiometric consistency across the two images. In addition, GPS devices were used to record field survey points, which were subsequently used to verify treeline positions and assist in manual ground truth labeling."

Reviewer: Regarding training, the authors mention using 512x512 patches and then splitting the dataset. Is the train/test split done at the patch level or at the pixel level (within patches)? This distinction is important, as pixel-level splits can introduce data leakage, especially in spatially autocorrelated datasets.

Response: Thank you for pointing this out. The dataset was split at the patch level, not the pixel level. We revised the manuscript as follows:

"Each image (5380 × 4671 pixels) was segmented into 110 non-overlapping patches of 512 × 512 pixels. The dataset split was performed at the patch level, not the pixel level, to avoid spatial autocorrelation and data leakage. Specifically, 80% of the patches were randomly selected for training and validation (75% for training, 25% for validation), and the remaining 20% of patches were used as the independent test set. The number of patches used for training, validation, and testing was 66, 22, and 22, respectively."

Reviewer: The use of Random Forest (RF) for variable importance analysis is questionable. This approach is valid only if variables are independent, which is clearly not the case here. Additionally, is it worth performing this complex selection to save 20% of variables? Reducing from 77 to 61 features may not justify the effort, especially if interpretability or performance gain is marginal. As such, the entire discussion about variable importance remains inconclusive.

Response: Thank you for your comment. We acknowledge that Random Forest (RF) variable importance measures can be biased when input features are correlated. In our dataset, some of the 77 features—such as vegetation indices and texture metrics—are derived from overlapping spectral bands and are therefore not entirely independent. Nevertheless, RF remains a widely used method for feature selection in high-dimensional remote sensing and ecological data, and its robustness has been demonstrated even in the presence of correlated variables (Cutler et al., 2007; Belgiu & Drăguț, 2016). In our study, RF was employed primarily to rank features and facilitate a conservative feature selection process, ultimately retaining the top 61 out of 77 features. This reduction resulted in a 14.3% decrease in training time and a slight improvement in overall accuracy (OA increased from 0.838 to 0.842). Although the gains were modest, this optimization was valuable considering the computational demands of the U-Net model.

Reference:

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J.: Random forests for classification in ecology. Ecology, 88(11), 2783-2792, https://doi.org/10.1890/07-0539.1, 2007

Belgiu, M., and Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens., 114, 24-31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Reviewer: Finally, the reported 14-meter height increase lacks context. The sentence "Forest area and highest point height difference from 2012 to 2021" is vague. Does this mean the authors extracted the maximum elevation value among all forest pixels? What was done to ensure robustness against outliers or noise? Also, scientific results are typically reported with associated uncertainties, which are missing here—or, if included, were not clear to me.

Response: Thanks for your comment. The 14-meter elevation gain was calculated based on the difference between the highest elevations of the forest area in 2021 and 2012. To elaborate on the finding, we have performed another analysis by investigating the elevation percentage of forest cover in 2021 and 2012. Based on the results, we revised the manuscript as follows:

> "Based on the 95th percentile of DEM elevation values of all pixels classified as forest (Fig. 9), the treeline showed an upward shift of 32.00 meters between 2012 and 2021. The 95% confidence interval (± 4.00 meters) was estimated using a bootstrap resampling method (5,000 iterations). Differences in area changes across various elevation ranges are detailed in Table 8, with the most significant changes occurring in the 3,500- to 3,600-m range. The most stable area was observed in the range of 3,700 to 3,800 m."

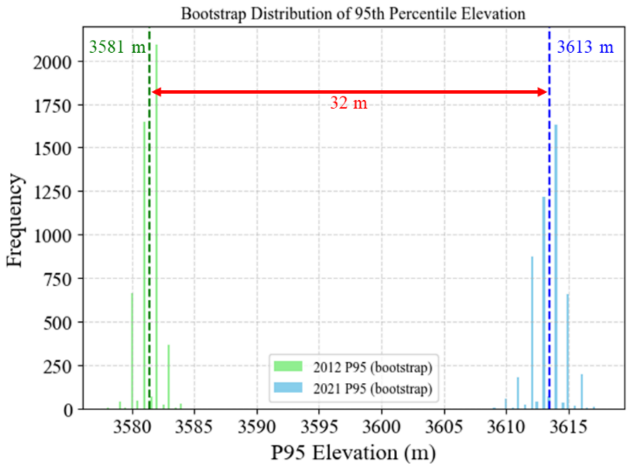

Figure 9. Bootstrap distribution of the 95th percentile elevation of forest cover for 2012 and 2021. The histogram shows the frequency of estimated 95th percentile elevations (P95) based on resampling. Green bars represent 2012 estimates, while blue bars represent 2021. The dashed vertical lines indicate the mean P95 value for each year.

Reviewer: Lastly, if the only interest was in changes to forest cover, why not classify the change directly instead of classifying each image independently?

Response: Thank you for the comment. We manually labeled each image. However, only the 2021 image was used for model training. The 2012 image was classified using the trained model, and the labeling for 2012 was done afterward for accuracy validation, not training. Direct change classification would require labels from both years or a different type of model. Our goal was to evaluate whether a model trained on recent data could still perform well on earlier imagery. Since both images were taken in autumn and we applied histogram matching, seasonal and lighting differences were minimized.

Minor Comments

"Taiwan has the highest density of high mountains globally, with over 200 peaks exceeding 3,000 meters in elevation."

→ This sounds too subjective. The result depends on the threshold chosen. I recommend rewriting as:

"Taiwan is one of the regions with the highest density of high mountains, with over 200 peaks exceeding 3,000 meters in elevation."

Response: Thank you for the comment. We revised the manuscript as follows:

> "Taiwan is one of the regions with the highest density of high mountains, with over 200 peaks exceeding 3,000 meters in elevation."

The introduction goes beyond the immediate scope of the study. However, I appreciate that the authors took the time to place their work in a broader context.

"At the same time, the productivity of alpine treeline vegetation increased, enhancing the ability to sequester atmospheric CO_2 and mitigating the effects of climate change (Rumpf et al., 2022)"

→ If this is true, however it's also be stated that the global effect is likely minor. The sentence could be more balanced.

Response: Thank you for the comment. We revised the manuscript as follows:

> "At the same time, the productivity of alpine treeline vegetation increased, enhancing the ability to sequester atmospheric CO_2 and mitigating the effects of climate change (Rumpf et al., 2022)"

Reference:

Rumpf, S. B., Gravey, M., Brönnimann, O., Luoto, M., Cianfrani, C., Mariethoz, G., and Guisan, A.: From white to green: Snow cover loss and increased vegetation productivity in the European Alps. Science, 376(6597), 1119-1122, https://doi.org/10.1126/science.abn6697, 2022.