Eya Cherif
Remote Sensing Centre for Earth
System Research (RSC4Earth)

Leipzig University
cherif@informatik.uni-leipzig.de

**To Biogeosciences**

27.08.2025

**Ref. No.: egusphere-2025-1284- "Uncertainty Assessment in Deep Learning-based Plant Trait Retrievals from Hyperspectral data"**

Dear Reviewer, dear Editor,

Thank you for your time reviewing our manuscript and for your constructive and helpful comments. In response to the comments and suggestions provided, the main changes to the manuscript will include:

1. **Methodology presentation:** We will improve the methodological section by adding equations describing the dissimilarity indices.
2. **Spectral data description and preprocessing**: We will add a dedicated supplementary section describing preprocessing in full detail, including band exclusions, interpolation methods, and smoothing parameters.
3. **Sensitivity analysis**: We conducted and will include a supplementary sensitivity analysis of Dis_UN across different quantile levels ($\tau$ = 0.75–0.99), to justify our choice of the 95th quantile.
4. **Novelty and Future Directions:** We extended the introduction and discussion to better distinguish Dis_UN from prior methods and to position it for future research in uncertainty quantification.

In addition to addressing the reviewer's comments, we have enriched the comparison to state-of-the-art uncertainty estimation methods. In the literature, both probabilistic and deterministic deep ensemble approaches are used; we now explicitly include both methods.

A detailed, point-by-point response, including the proposed changes in the manuscript, are attached.

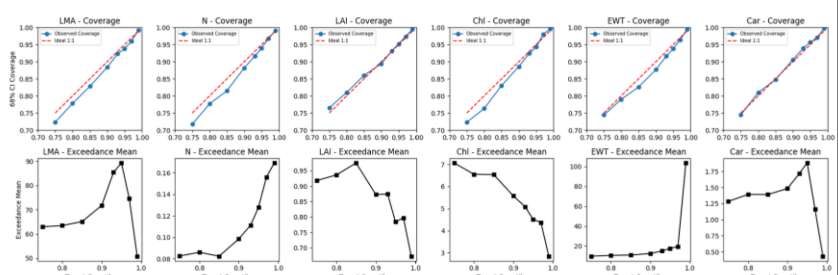
Kind regards,

Eya Cherif (on behalf of all co-authors).

<table>
<tr><th colspan="3">Responses to individual comments from RC2</th></tr>
<tr><td>ID</td><td>Comment</td><td>Response</td></tr>
<tr><td></td><td>This study introduces a distance-based uncertainty quantification method (Dis_UN) that improves the reliability of plant trait retrievals from hyperspectral data, particularly in out-of-domain (OOD) scenarios. This represents an advancement for robust vegetation monitoring and ecological applications. However, there are still a few issues that need to be considered as follows.</td><td>We would like to thank the reviewer for these very constructive comments that were very helpful to further improve the manuscript. Please find below a detailed answer to the comments.</td></tr>
<tr><td>1</td><td>The Novelty and Comparisons with Existing Work: The manuscript highlights the limitations of traditional uncertainty quantification methods (Ens_UN and MCdrop_UN) in OOD scenarios , and positions Dis_UN as a solution to these challenges. To further strengthen the claim of novelty, it is suggested to provide a more detailed discussion on how Dis_UN specifically differentiates itself from other distance-based uncertainty methods mentioned (e.g., Silvan-Cardenas et al., 2008; Khatami et al., 2017; Feilhauer et al., 2021). A brief table summarizing key differences could be considered.</td><td>We thank the reviewer for this comment. We would like to clarify that the cited studies (Silván-Cárdenas and Wang, 2008; Khatami et al., 2017; Feilhauer et al., 2021) have made important contributions by using distance- or probability-based metrics to quantify uncertainty in classification tasks, particularly in the context of land cover or vegetation mapping. However, these approaches are not directly comparable to our study, which focuses on a regression setting for trait estimation using deep learning models. To ensure comparability, we benchmarked our method against established state-of-the-art uncertainty quantification techniques commonly used in deep learning, such as deep ensembles and Monte Carlo dropout.<br><br>Rather than providing a comparison table, we will add clarifying sentences in the introduction to better articulate these distinctions and to position our method relative to previous work. We hope this provides the necessary context to highlight the novel contribution of the Dis_UN approach.<br><br>“Introduction<br><br>…Instead of building on the variance in the predictions, uncertainty estimation for EO should particularly focus on the dissimilarity between the training and the new data. In other words, if an observation is very different from what the model has learned, it is likely to be very uncertain (Meyer and Pebesma, 2021, Linnenbrink et al., 2024). There is, therefore, a need for an uncertainty estimation approach that accounts for dissimilarities between the training and unseen data (Silvan-Cardenas et al., 2008; Khatami et al., 2017; Feilhauer et al., 2021).<br><br>To address the challenges of uncertainty quantification, especially in OOD scenarios, distance-based methods have emerged as a promising solution.</td></tr>
</table>

| | | |
|---|---|---|
| | | *Earlier work has applied similarity-based metrics in the context of classification (Silvan-Cardenas et al., 2008; Khatami et al., 2017; Feilhauer et al., 2021). These approaches remain tied to discrete categorical problems and shallow empirical models. More recently, distance-based methods have been extended to regression and spatial prediction tasks. For instance, Janet et al. (2019) proposed a low-cost uncertainty metric for predictions of chemical properties of unknown substances/materials based on the distance of new inputs from the training data in latent space, outperforming traditional uncertainty metrics such as Monte-carlo dropout and ensembles, particularly for data points far from the training set. Meyer and Pebesma (2021) discussed the importance of defining an "area of applicability" for spatial models, emphasizing the use of dissimilarity metrics to assess model confidence when dealing with new data. Building on this, Papacharalampous et al. (2024) and Linnenbrink et al. (2024) illustrated the effectiveness of distance-based metrics in enhancing uncertainty quantification and improving model reliability in spatial predictions."* |
| 2 | Clarification of Ecological Applications: While the manuscript lists several ecological applications (e.g., biodiversity monitoring, Earth system modeling, vegetation health assessment), it is suggested to clarify which specific aspects of these applications benefit most from reliable uncertainty quantification. For instance, this could involve identifying areas where model predictions are less trustworthy, or guiding more targeted field campaigns. | We thank the reviewer for this comment. In response to the suggestion, we will further expand the introduction to include the importance of uncertainty propagation in downstream modeling workflows.<br><br>*"**Introduction***<br><br>*…While some efforts have been made to quantify uncertainty in the context of hyperspectral plant trait retrieval (García-Soria et al., 2024; Singh et al., 2015; Wang et al., 2019), the results are often not comparable as the definition and interpretability of the uncertainty estimates varies depending on the methods used. Uncertainty quantification is particularly prevalent in EO for vegetation monitoring, where training data is typically sparse, and models are often applied to new, unseen regions and, hence, data that are OOD (Kattenborn et al., 2022; Ploton et al., 2020, Meyer and Pebesma 2021). In addition to providing crucial information on the quality of OOD predictions, quantitative estimates of uncertainty are increasingly utilized in a range of downstream ecological and environmental applications and are often required by data assimilation schemes in order to appropriately weigh all available observations (Chernetskiy et al., 2017; Lewis et al., 2012; Mathieu and O'Niell, 2008). In applications such as the assessment of land surface phenology, for example, recent studies have explored the propagation of plant trait prediction uncertainties to derived phenological metrics (Graf et al., 2023), enabling more robust detection of changes in phenophases (e.g., due to the effects of climate change). In ecological modeling more broadly, incorporating trait-level uncertainty allows for realistic error propagation, thereby increasing the robustness of simulations related to vegetation dynamics, biodiversity assessments, and Earth system forecasts. Uncertainty estimates also serve as a valuable tool for identifying underrepresented conditions in the training set. By highlighting regions with high uncertainty,* |

| | | |
|---|---|---|
| | | *they can inform active learning strategies and guide targeted data acquisition campaigns, ultimately improving both model generalization and data representativeness. The increasing importance of uncertainty estimates is reflected by the recent efforts of space agencies and data providers (Brown et al., 2021b; Gorroño et al., 2018, 2017; Goryl et al., 2023), and uncertainties are now a goal of Analysis Ready Data (ARD) standards (Committee on Earth Observation Satellites, 2024, https://ceos.org/ard/). In addition, recent reports from the European Commission (Camia et al., 2024), highlight uncertainty estimations as a specific quantitative requirement for various European Union land-related environmental and agriculture policies."* |
| 3 | Mathematical Formulation for Dissimilarity Index: To enhance the self-contained clarity of the methodology, please include the mathematical formulation for the dissimilarity index (DI) directly within Section 2.1.2. | We thank the reviewer for this suggestion, which helped improve the clarity and readability of our methodology. In response, we will revise Section 2.1.2 by including a mathematical formulation of the dissimilarity index (DI), which was used as a core predictor in our uncertainty estimation approach. The added equations will describe the cosine distance calculation between test and training spectra, the procedure for summarizing distances via the median of the 50 nearest neighbors, and the normalization against the training set mean (Equations 1–3).<br><br>We will add the corresponding clarification under the 2.1.2 Dissimilarity indices (predictors) section:<br><br>*"**2.1.2 Dissimilarity indices (predictors)***<br><br>*The DI, used as a predictor in this study, was calculated using the cosine distance, a well-suited metric for analyzing reflectance data. The cosine distance effectively captures the angular relationship between two spectra (Kruse et al., 1993), emphasizing spectral shape while minimizing the influence of amplitude variations that occur uniformly across the spectrum. This helps mitigate brightness changes caused by heterogeneous illumination and internal shading (Feilhauer et al. 2010).*<br><br>*Formally, the cosine distance between a test spectrum xi and a training spectrum zi is defined as:*<br><br>$CosineDist(x_i, z_j) = 1 - \frac{x_i . z_j}{\|x_i\| . \|z_j\|}$ *(1)*<br><br>*This DI was applied in both the feature space and the embedding space of the models (Fig. S2). As a first step, we calculated cosine distances between each sample of the test dataset xi and the samples of the training data set zi. These calculations were performed using the Python package FAISS (Douze et al., 2024), which is optimized for fast similarity search and clustering of large datasets. As a next step, each DI was calculated as the median of the distance distribution between a test sample and its 50 nearest neighbors in the training set:* |

| | | |
|---|---|---|
| | | $DI_i = median\{CosineDist(x_i, z_j)\}_{j=1}^{50} \quad (2)$<br><br>*To ensure comparability across samples, the indices were normalized against the mean DI value of the entire training set (Meyer and Pebesma, 2021):*<br><br>$DI_i^{norm} = \frac{DI_i}{\mu_{train}}$ *, with* $\mu_{train} = \frac{1}{n}\sum_{j=1}^{n} DI_i$ *where n is the number of training samples* *(3)* " |
| 4 | Sensitivity Analysis for 95-Quantile Regression: The choice of 95-quantile regression is justified as estimating worst-case errors. However, please include a brief discussion or a supplementary analysis on the sensitivity of Dis_UN's performance to this specific quantile choice (e.g., how results might differ with 90th or 99th percentile). This would strengthen the methodological rigor. | We thank the reviewer for this suggestion. In response, we conducted a sensitivity analysis to evaluate the empirical coverage obtained using Dis_UN models trained at various quantile levels (τ), ranging from the 75th to the 99th quantile for all six target variables. For each τ, we computed the empirical coverage (with 68% Wilson CIs), and exceedance severity (mean magnitude of residuals exceeding the predicted bound). The results, which will be included in the Supplementary Material (Figure below) show that for all six traits, observed coverage increases nearly on the 1:1 line with the target quantile, reaching ≈0.94–0.96 at τ=0.95. The exceedance mean varies smoothly up to τ=0.95 but shows abrupt jumps at τ=0.97–0.99 (e.g., EWT and Car), indicating entry into a less stable, overly conservative tail regime. These results support τ=0.95 as a data-driven compromise that attains high coverage while avoiding the sudden instability seen at extreme quantiles.<br><br>We will additionally add a short paragraph to justify our choice in the method section:<br><br>***"2.1.3 Dis_UN Model Training***<br><br>***…****To further support the choice of the 95th quantile, we conducted a sensitivity analysis across a range of quantiles for all traits (Fig. below). The results show that τ = 0.95 provides a good balance between capturing a high proportion of large errors and avoiding overly wide and unstable uncertainty bounds. Lower quantiles (τ ≤ 0.93) tend to underestimate the extent of potential errors, missing a fraction of extreme cases, while higher quantiles (τ ≥ 0.97) lead to unnecessarily conservative bounds that can fluctuate sharply. This balance makes the 95th quantile a robust choice for representing worst-case uncertainty across variables, avoiding both underestimation and over-conservatism."*<br><br> |

| | | |
|---|---|---|
| 5 | Dataset Diversity and Bias Handling: The manuscript describes a "heterogeneous training set" compiled from 50 datasets across various ecosystems , but also acknowledges inherent biases and a lack of fully global representation. Please expand on the specific strategies employed during dataset curation to minimize known sampling biases across regions, species, and biomes. Even if complete mitigation was not possible, detailing the efforts made would be beneficial. Please also explicitly refer to Tables S2 and S3 in the main text when discussing data sources and their representativeness. | We agree with the reviewer and have clarified the bias-handling steps in the Methods. Specifically, the imbalance was reduced in two stages:<br><br>1. **During training of the multi-trait CNN**, we applied random upsampling with replacement so that each dataset contributed equally per epoch, and used a weighted loss function where samples from under-represented datasets received larger weights (cherif et al. 2023).<br><br>2. **During training of Dis_UN**, we applied a per-sample loss weights inversely proportional to the number of samples, which down-weights over-represented residual ranges and up-weights under-represented ones for the quantile regression (see Section 2.1.3)..<br><br>We will add a dedicated clarification in section 2.1.1, and refer to Tables S2 and S3 when discussing the sources and representativeness of the training data.<br><br>**"2.1.1 The Multi-trait Model**<br><br>***…****Yet, the collected data do not provide a fully global representation due to the labor-intensive nature of data collection and the inherent bias in available data measurements (Table S2). To reduce over-representation of large datasets during training of the multi-trait CNN, we (i) performed random upsampling with replacement so that each source dataset contributed approximately equally per training epoch, and (ii) applied per-sample loss weights inversely proportional to the number of labeled samples in the corresponding source dataset, which down-weights over-represented datasets and up-weights under-represented ones."* |
| 6 | Strategies for Spectral Saturation Challenges: The manuscript identifies spectral saturation as a remaining challenge for certain traits. In the discussion or future work section, please propose specific experimental avenues or mitigation strategies to address this limitation for affected traits. For example, exploring alternative spectral indices, radiative transfer models, or machine learning | We appreciate the reviewer's point; spectral saturation of traits like LAI is a long-standing and physics-governed challenge (Zheng et al. 2009, Camps-Valls et al. 2021). However, this limitation is not methodological but physical, arising from radiative transfer: once foliage coverage becomes high, additional leaf layers contribute little new information to top-of-canopy reflectance. The fundamental problem is that we cannot see through leaves, so lower leaf layers may not alter the spectral responses. This saturation is well documented for NDVI and LAI (e.g., Sellers, 1985; Myneni et al., 2002; Gitelson, 2004, Steltzer and Welker. 2006, Xu et al. 2020) and persists across spectral indices, radiative transfer models, and machine learning methods. Thus, the saturation problem reflects the physics of light–canopy interaction rather than shortcomings of the specific algorithms we use.<br><br>In the manuscript (Outlook), we will cite these studies, emphasize the physical basis of saturation, and acknowledge it as an open, long-standing limitation. We also note that mitigations may be possible, such as incorporating spatial |

| | | |
|---|---|---|
| | architectures less susceptible to saturation. | context, or exploiting multi-angular data, but that no optical method can fully overcome the physics-driven saturation.<br><br>*"**4.5 Outlook: Uncertainty in the Context of Global Trait Mapping***<br><br>***…**At the same time, it is important to recognize that distance-based uncertainty estimation cannot by itself overcome data-intrinsic limitations. Structural traits such as leaf area index are affected by the long-recognized problem of spectral saturation, where top-of-canopy reflectance becomes insensitive to additional foliage at high canopy densities (e.g., Sellers, 1985; Myneni et al., 2002; Gitelson, 2004, Steltzer and Welker. 2006, Zheng et al. 2009, Xu et al. 2020). In such cases, saturation arises from the inherent distribution of the data and constrains both training and inference. Distance-based uncertainty is therefore best understood as a diagnostic tool that reveals these information gaps, rather than as a mechanism to eliminate them. Progress will require more sophisticated sensing strategies, where recent work has shown promising directions (Mutanga et al., 2022; Wan et al., 2022). However, no purely optical method fully overcomes saturation, as this limitation is rooted in the physics of canopy reflectance. This limitation, in turn, motivates the continued development of distance-aware uncertainty methods that more explicitly link the training samples to unseen data. Importantly, such methods are not restricted to vegetation trait retrieval but can be readily applied and extended to a wide range of applications where robustness under distribution shift and reliable uncertainty quantification are critical.*<br><br>*References*<br><br>*Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., ... & Running, S. W. (2021). A unified vegetation index for quantifying the terrestrial biosphere. Science Advances, 7(9), eabc7447.*<br><br>*Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of plant physiology, 161(2), 165-173.*<br><br>*Mutanga, O., Masenyama, A., & Sibanda, M. (2023). Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. ISPRS Journal of Photogrammetry and Remote Sensing, 198, 297-309.*<br><br>*Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. IEEE Transactions on Geoscience and remote Sensing, 33(2), 481-486.* |

| | | |
|---|---|---|
| | | *Sellers, P. J. (1985). Canopy reflectance, photosynthesis and transpiration. International journal of remote sensing, 6(8), 1335-1372.*<br><br>*Steltzer, H., & Welker, J. M. (2006). Modeling the effect of photosynthetic vegetation properties on the NDVI–LAI relationship. Ecology, 87(11), 2765-2772.*<br><br>*Wang, Q., Putri, N. A., Gan, Y., & Song, G. (2022). Combining both spectral and textural indices for alleviating saturation problem in forest LAI estimation using Sentinel-2 data. Geocarto International, 37(25), 10511-10531.*<br><br>*Xu, D., An, D., & Guo, X. (2020). The impact of non-photosynthetic vegetation on LAI estimation by NDVI in mixed grassland. Remote Sensing, 12(12), 1979.*<br><br>*Zheng, G., & Moskal, L. M. (2009). Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. Sensors, 9(4), 2719-2745."* |
| 7 | Additional Potential Limitations: Consider briefly discussing the computational cost associated with the training phase of the Dis_UN model itself. While inference is computationally efficient, the training cost could be a factor for extremely large datasets. | We appreciate the reviewer's point. In response to this comment we will add clarification in the discussion to distinguish between the training and inference phase running time and we will add and refer to the results in the supplemental materials:<br><br>The modified version will read as follows:<br><br>**"4.2.1 Comparison with other methods**<br><br>…*Beyond enhancing uncertainty prediction performances, the proposed distance-based uncertainty estimation method provides substantial computational advantages over variance-based approaches. Unlike variance-based methods that require multiple forward passes to compute prediction variance, the distance-based approach allows for straightforward application once the uncertainty model is trained. This eliminates the need for repeated inference runs, making it significantly more computationally efficient (Table S6 below). Such efficiency is particularly valuable for large-scale remote sensing applications, where fast and scalable uncertainty estimation is crucial.*<br><br>*Though, it is important to distinguish between the training and inference costs of the proposed method (Table S6 and S7). In our experimental setup, the training time of the distance-based uncertainty model was of a similar order to that of deep ensembles, as we adopted a leave-one-dataset-out (LODO) transferability analysis to explicitly evaluate out-of-distribution conditions, requiring the training of 50 models. However, this design reflects a specific validation strategy rather than an intrinsic requirement of the approach. In practice, distance-based uncertainty estimation can be integrated into more conventional validation schemes, such as k-fold cross-validation, thereby substantially reducing the training overhead."* |

| Training | Transferability training | Distance calculation | Training |
|---|---|---|---|
| Dis_UN | ~6h*50 | 0.34 h | ~1s |
| Ens_prob_Un | ~10h*50 | - | - |
| Ens_prob_Un | ~6h*50 | - | - |
| MCdrop_UN | ~6h | - | - |

**Table S7** Running time of the benchmarked methods for uncertainty estimation during the training phase

| Inference | Ens_prob_Un | | Ens_det_Un | | MCdrop_UN | | Dis_UN | |
|---|---|---|---|---|---|---|---|---|
| | Enmap | NEON | Enmap | NEON | Enmap | NEON | Enmap | NEON |
| Data preparation | 6.64 | 9.62 | 6.64 | 9.62 | 6.64 | 9.62 | 190.9516 | 203.877 |
| Model application | 486.8539 | 556.7541 | 483.3784 | 555.3588 | 1567.305 | 1774.845 | 0.0889 | 0.0746 |
| Total | 493.4939 | 566.3741 | 490.0184 | 564.9788 | 1573.945 | 1784.465 | **191.0405** | **203.9516** |

**Table S8** Running time in seconds of the benchmarked methods for uncertainty estimation during the inference phase

| | | |
|---|---|---|
| 8 | Abstract Terminology Simplification: While terms are defined later in the manuscript, consider simplifying technical jargon in the abstract, such as "predictor and embedding space", for broader accessibility. Phrases like "data characteristics and model features" might be more immediately understandable. | We thank the reviewer for pointing this out and agree that the abstract should be accessible to a broad readership. Here, we have to particularly serve the ecology, remote sensing and data science communities. Thus, in the revised abstract, we retained the technical terms for precision but added short clarifications in parentheses—"predictor space (spectral data )" and "embedding space (features learned by the deep model)"—so that the terminology is both accessible to the different communities.<br><br>*"**Abstract:***<br><br>*…To address this limitation, we propose a distance-based uncertainty estimation method (Dis_UN) that quantifies prediction uncertainty by measuring dissimilarity in the predictor space (spectral data) and embedding space (features learned by the deep model) between training and test data. Dis_UN leverages residuals as a proxy for uncertainty and employs dissimilarity indices in data manifolds to estimate worst-case errors via 95-quantile regression.2"* |
| 9 | Figure Caption Enhancement: Figure captions (e.g., Figure 3 , Figure 4 , Figure 5 , Figure 6 , Figure 7 ) could be more descriptive. For instance, for Figure 3, explicitly state what the X and Y axes represent (e.g., "Predicted Uncertainty vs. Observed Residuals"). For Figures 4, 5, 6, and 7, please ensure the captions clearly explain the relationship between the spatial maps, the box/violin plots, and the JM/KS values to guide the reader through | We will check and improve the description of the figures' captions. |

| | | |
|---|---|---|
| | the interpretation of uncertainty comparisons. | |
| 10 | Early Definition of OOD Data: Although the concept of OOD data is contextualized well (Lines 29-30, 48-53), for immediate comprehension, please provide a concise and explicit definition of "OOD data" with concrete examples (e.g., "unseen geographical regions, species, biomes, different sensors, or scene components like clouds and water bodies") at its first introduction. | We thank the reviewer for this suggestion. We revised the abstract (and ensured consistency with the introduction) to include an immediate definition of OOD data. It now reads:<br><br>*"**Abstract:***<br><br>*…out-of-distribution (OOD) data, i.e. test samples that differ substantially from the training distribution, such as unseen geographical regions, species, biomes, sensors, or scene components (e.g., clouds, water bodies)."* |
| 11 | Writing Flow and Conciseness: The manuscript is generally well-organized and readable. However, some sentences could be rephrased for improved conciseness or flow. For example, the sentence describing the two phases of the method (Lines 129-130) could be streamlined for better readability. | We will carefully check and improve the readability and flow of the entire manuscript in the revised version. |
| 12 | Detailed Preprocessing in Supplementary Material: While some preprocessing steps are mentioned (Lines 145-147), for full reproducibility, please include a dedicated supplementary section with more detailed preprocessing steps, including specific interpolation methods, filtering parameters, and | In response to the reviewer's suggestion, we will add a supplementary section adding detailed information for our data processing. The section will read as follows:<br><br>***"S1. Preprocessing pipeline***<br>*All 50 compiled datasets were pre-processed using the same standardized pipeline, without any dataset-specific deviations. The procedure followed here is based on the analysis of Cherif et al. (2023) and summarized as follows:*<br><br>*First, reflectance spectra were quality-checked. Reflectance values outside the physical range were masked: values below zero were set to missing, and* |

| | | |
|---|---|---|
| | precise band exclusions for each dataset. | *values greater than one were treated as spurious spikes. These missing values were then replaced by the mean of the nearest valid neighbors.*<br><br>*Second, all datasets were resampled to a common 1 nm resolution to harmonize the diverse measurements from different sensors (from proximal and airborne), which varied in spectral sampling and band centers. This resampling was not intended to enhance spectral resolution or recover fine-scale features absent in coarser sensors, but to provide a uniform input representation required for the deep learning model. Most datasets were already acquired at 1 nm resolution, so upsampling was preferred over downsampling to minimize manipulation of the data and avoid loss of information from higher-resolution sensors.*<br><br>*Third, spectral intervals strongly affected by atmospheric water absorption were excluded uniformly across all datasets. In the implementation, the following wavelength ranges were removed: 1351–1430 nm, 1801–2050 nm, and 2451–2500 nm.*<br><br>*Finally, the remaining reflectance data were smoothed using a Savitzky–Golay filter applied independently to three contiguous segments of the spectrum: 400–1350 nm, 1431–1800 nm, and 2051–2451 nm. Each segment was filtered with a window size of 65 nm and a polynomial order of one."* |
| **13** | Data Availability Clarification: The GitHub link for code availability (Line 1043) is excellent. Please explicitly state whether the full processed datasets used for training and testing are also available or how they can be accessed (e.g., if too large for GitHub, mention a data repository). | We thank the reviewer for this recommendation. The training datasets of the deep learning model were made available as part of the previous study (cherif et al. 2023). However, we will share the distance data relevant for the current study. Due to the large size of the data, we will share it via other platforms like HuggingFace. |
| 14 | Applications in Other Remote Sensing Domains: Expand the discussion on potential applications of Dis_UN beyond vegetation monitoring. For instance, discuss how it could be applied to uncertainty quantification in other remote sensing domains, such as land cover | We thank the reviewer for this comment. In response we will add the following to the outlook section.<br><br>*"**4.5 Outlook: Uncertainty in the Context of Global Trait Mapping***<br><br>*..Importantly, such methods are not restricted to vegetation trait retrieval but can be readily applied and extended to a wide range of applications where robustness under distribution shift and reliable uncertainty quantification are critical."* |

| | | |
|---|---|---|
| | classification, deforestation detection, or urban mapping, where OOD conditions (e.g., new urban structures, novel land cover types) are common. | |
| 15 | Integration with Other Data Modalities: In the future work section, suggest avenues for integrating Dis_UN with other data modalities beyond hyperspectral (e.g., LiDAR or SAR data) for a more comprehensive uncertainty assessment, particularly for structural traits. This would further enhance the model's generalizability and impact. | We thank the reviewer for this suggestion. The proposed method is inherently sensor-agnostic and can be applied across different remote sensing modalities or in multi-sensor settings. Modalities such as LiDAR, SAR, or high-resolution canopy height models are indeed valuable for improving trait predictions and thereby indirectly reducing predictive uncertainty. However, the aim of this study is to develop and test a post-hoc framework for uncertainty quantification, independent of the input data and based solely on how the training and test samples relate to each other. |