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

Leipzig University
eya.cherif@uni-leipzig.de

To Biogeosciences
16.12.2025

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

Dear Reviewers, dear Editor,

Thank you for your time reviewing our manuscript and for your positive feedback. In response to the minor comments, the main changes to the manuscript will include:

1. **Fair calibration of variance-based methods:** All ensemble- and dropout-based uncertainties are now consistently scaled to the 95% confidence level (1.96 × σ). Figures 3, 6, and 7 and the corresponding discussion were updated accordingly to ensure a fair comparison with the Dis_UN quantile-based uncertainty.
2. **Highlight the advantages of Dis_UN:** The Discussion and Conclusion were expanded to explicitly state that while ensemble methods show a good average-case calibration for OOD vegetation, Dis_UN uniquely provides a conservative, assumption-free upper error bound and superior landscape-scale OOD contrast, between vegetated and non-vegetated pixels particularly.
3. **OOD distribution contrast metric updated:** we replaced the Jeffries–Matusita (JM) distance with the Kolmogorov–Smirnov (KS) distance as the primary measure of distribution contrast throughout the landscape-scale OOD analysis. JM tends to saturate rapidly therefore loses sensitivity to differences in strong OOD.

A detailed, point-by-point response, including the proposed changes in the manuscript, can be found below.

Kind regards,

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

Responses to individual comments from RC1		
ID	Comment	Response
	GENERAL COMMENTS	
	I thank the authors for the careful review and complete answer to comments. I think the manuscript has overall improved, and most methodological and analytical questions have been clarified. While for the most part, the manuscript should be ready for publication, I still have concerns regarding the comparison of the different uncertainty metrics. As exposed below, I still think that the comparison of metrics is biased by construction. A fair comparison will still prove the advantage of the new method proposed by the authors and reveal it more accurately.	We thank the reviewer for their careful and thoughtful evaluation of our manuscript and for the detailed clarification regarding probability coverage in the comparison of uncertainty estimates. We appreciate the reviewer's constructive reasoning and now fully see the point raised. In the original submission, our rationale for presenting the ensemble and dropout uncertainties in their standard form was to transparently illustrate the extent to which these methods underestimate uncertainty relative to their corresponding observed residuals, without any post-hoc scaling. At that stage, our intention was not to compare the absolute magnitudes of ensemble uncertainties directly against Dis_UN, but rather to show the widely documented underestimation trend as it appears when these methods are applied "as is." However, we acknowledge the reviewer's concern that presenting the methods in their unscaled form may introduce confusion for readers and could hinder a fair, side-by-side comparison. We agree that matching the confidence level across methods provides a clearer and more rigorous evaluation. Following the reviewer's recommendation, we have now implemented all requested changes: 1. Figures 3, 6, and 7 have been replaced with the calibrated variance-based methods' (95% confidence interval) versions. 2. The Results and Discussion sections have been revised to reflect these recalibrated comparisons. 3. All supplementary material has been updated to ensure consistency across the manuscript.
	SPECIFIC COMMENTS	
1	I have read with attention the response to the previous comment on the differences in coverage between the uncertainty metrics compared. I thank the authors for addressing this question and preparing Fig. S6. The answer, however, has	We thank the reviewer again for helping us strengthen the fairness, clarity, and conceptual rigor of the manuscript.

	further convinced me that the comparison of the different uncertainty metrics is biased by construction. While the authors argue that the approaches are not conceptually comparable, it could be argued that they can be as long as the assumption of normal distribution is met (which does not necessarily have to happen), and that the choice of reporting the model uncertainty as the standard deviation, to some extent, accepts this assumption. In practice, applying the 95% coverage to the ensemble methods makes the ensemble approaches closer to the results of the new method proposed in this manuscript. The results of the ensemble methods in Fig. S6 are much more comparable to what is presented in Fig. 3, and maybe, in some cases, even better. The fact that the methods are applied "as they are typically applied in the literature" is not a robust argument; these methods could be applied as such because, in the "standard uncertainty" shape, the reported uncertainty can easily be expanded to the desired confidence level (under the assumption of normality). Thus, the way uncertainty is reported is not meant to represent the upper bound of the absolute error, but the conversion to 95 % is straightforward. Thus, I am not convinced by the authors that the comparison in Fig. 3 is fair. If error distributions are normal, the 95 % bounds of the absolute error would be equivalent for both approaches, or very similar, wouldn't they? If so, comparing the standard error will lead to a systematic difference (i.e., underestimation) because the probability covered is different (i.e., smaller). The question then is whether, when covering the same probability, the uncertainty estimates of the state-of-the-art methods are worse, comparable, or better than the new approach proposed. Therefore, I still think Fig. S6 should be presented instead of Fig. 3. The same conversion (multiplying by 1.96) should be applied to Fig. 6 and 7; in this case, the advantage of the new method proposed would be clearer. Benchmarking this way does not demerit the approach proposed; however, it acknowledges that the difference with other methods might be lower than the one depicted when certain	*The changed results section now reads:* ***"3.1 Uncertainty for OOD Vegetation data*** *To ensure a consistent comparison between Dis_UN and the variance-based approaches, we scaled the ensemble- and dropout-based uncertainties to the 95% confidence interval, i.e 1.96 × σ (Fig. 3). Results with the default scale, i.e 1 × σ as commonly used, are presented in Fig. S6 in the Supplementary. The MCdrop_UN method showed the strongest underestimation of residuals, independent if the uncertainty estimate was scaled or not and only covered 3.1% to 7.9% (QuRatio) of the observed residual variability (Fig. 3, Table. S5). This reflects a very narrow predicted interval that does not correspond well to the actual errors. The ENCE values, used to quantify the uncertainty prediction calibration, were the highest among all methods (13.1–20.8), indicating a weak relation between actual residuals and predicted uncertainty (Table S4).* *The two ensemble-based methods exhibited better calibration only after scaling the predicted uncertainty (1.96 × σ). Ens_det_UN achieved the closest alignment with observed residuals among the variance-based methods, with ENCE values between 0.12 and 0.65 and QuRatio between 43.2% to 60.1%. Note that without scaling, which is the common approach in literature, the ensemble approaches greatly underestimate the range of the observed residuals (Fig. S6)." (Lines 390-401)* *The updated discussion (Sections 4.1) "**4.1 Local-Scale Uncertainty in OOD Vegetation data*** *Applying the ensemble and Monte Carlo methods with their default settings resulted in a critical underestimation of observed residuals (Fig. S6). After scaling variance-based uncertainties to approximate 95% confidence intervals (1.96 × σ), the ensemble approaches showed substantially improved alignment with observed residuals (Fig. 3), in contrast to their unscaled performance shown in Fig. S6. The improved performance of both ensemble methods for OOD vegetation samples indicates the average alignment under natural distribution shifts (Gustafsson et al., 2020), that is, when new data differ in source or environmental conditions but still represent the same underlying object class (vegetation). However, these methods do not*

	conditions (normality and the implicit symmetry) are met. The strength of the proposed method lies then in the cases where such conditions are violated, and the fact that these do not need to be checked to determine whether the uncertainty estimates are to be trusted. This new method would be advantageous in multiple situations (e.g., the NEON and EnMAP imagery shown), and its use might be more recommendable for automated and large-scale processing.	*automatically adapt to distributional changes if not validated with an independent set.* *This limitation highlights the need for post-hoc scaling and introduces practical challenges in operational settings: practitioners must determine appropriate scaling factors for each trait and application context, requiring either assumptions of normality or empirical calibration on held-out data, which may not be available for novel sensors or ecosystems. Recent studies have increasingly emphasized the need for robust recalibration strategies for variance-based uncertainty methods, particularly under distributional shift (Liu et al., 2021; Ovadia et al., 2019; Palmer et al., 2022). While a comprehensive evaluation of such calibration strategies lies beyond the scope of this study, our results underscore that naïve application of variance-based methods without careful consideration of error distribution characteristics can lead to systematically biased uncertainty estimates.* *However, the Monte Carlo approach continued to perform poorly (high ENCE values) even after scaling, indicating its failure to capture variability within samples from different vegetation types. The observed low alignment between predicted uncertainty and residuals suggests that the uncertainty estimates produced by these models do not fully represent the model's errors (Fig. 3). This underestimation, especially for higher predicted values, is a known limitation of Monte Carlo-based approaches (Hu et al., 2022; Klotz et al., 2022; Liu at al. 2021), which tend to be optimistic in their uncertainty estimates. (Lines 525-544)"* *Caption of the updated figure. 3:* ***"Scatter plots comparing the predicted uncertainties (x-axis) from four methods— Distance-based (Dis_UN), probabilistic ensemble (Ens_prob_UN), deterministic ensemble (Ens_det_UN) and Monte Carlo dropout (MCdrop_UN) calibrated by a factor of 1.96 × standard deviation—against observed residuals (y-axis) of the multi-trait models across six traits: Leaf Mass per Area (LMA), Nitrogen content, Chlorophyll content (Chl), Equivalent Water Thickness (EWT), Leaf Area Index (LAI), and Carotenoid content (Car). Each point represents a sample, colored by vegetation type. For each trait–method combination, the regression line***

		(blue) is compared to the 1:1 line (black) to visualize alignment between predicted and observed errors. Model calibration is quantified by the Expected Normalized Calibration Error (ENCE), while the ratio of predicted to observed uncertainty ranges (QuRatio) indicates the coverage of residual variability (see also Table S4 and S5). Note that the default definition of uncertainty with the Ensemble and Drop out approaches did systematically underestimate the observed residuals (Fig. S6). *"*
	TECHNICAL CORRECTIONS	
	Lines 43-44: EWT is a leaf trait, maybe, place it before LAI, so that foliar and structural traits are packed together.	We have changed the text accordingly.
	Lines 140: Plant traits symbology has already been defined before, and is no longer necessary	We have changed the text accordingly.
	Line 204: The term "resolution" is misused. Likely, the authors refer to the "spectral sampling interval", not "**spectral resolution**". See, for example, Fig. 2 in Porcar-Castell et al. 2015 (https://doi.org/10.5194/bg-12-6103-2015/).	We have changed the text accordingly. *"In line with Cherif et al. (2023) (S1), all datasets were resampled to a common 1 nm spectral interval across the 400–2500 nm range to harmonize diverse measurements. We chose to upsample rather than downsample as most datasets were originally acquired at a 1 nm spectral sampling interval, thereby minimizing data manipulation. (Lines 182-185)"*
	Line 324: Set "k" to underscript format.	We have changed the text accordingly.
	Figure S6. Something seems to have gone wrong for Nitrogen in Ens_prob_UN. Were the errors multiplied by 19.6 instead of 1.96?	Thank you for pointing this out. This has now been corrected.

Responses to individual comments from RC2		
ID	Comment	Response
	The authors have addressed the reviews, which improves the manuscript. Revisions to methodology (e.g., DI formulas, flowchart), data processing (resampling), and discussion (LAI saturation) have enhanced the clarity and rigor. The main contribution—the Dis_UN method's effectiveness for OOD uncertainty—is now supported. The manuscript is nearly ready for publication, pending a few minor suggestions.	We would like to thank the reviewer for this positive feedback. Please find below a detailed answer to the minor comments.
1	Supplement with a Direct Assessment of the Dis_UN Model's Own Calibration. The manuscript currently focuses on the ENCE (Expected Normalized Calibration Error) and the correlation between uncertainty predictions and residuals (Fig. 3). It is suggested to add a more direct validation of the 95% quantile regression model itself: its "Empirical Coverage." Specifically, what percentage of the observed residuals in the LODO-CV test sets actually fall below the model's predicted 95% quantile bound? This would provide a more direct proof of the model's own calibration performance.	We thank the reviewer for this suggestion. We have already included a comprehensive empirical coverage analysis in our supplementary materials (Fig. S2). This figure also shows the empirical coverage of Dis_UN across quantile levels τ=0.75 to τ=0.99 for all six traits, with 68% Wilson confidence intervals. The results demonstrate that at τ=0.95 (our chosen quantile), empirical coverage reaches approximately 94-96% across all traits, closely matching the target 95% coverage. To better highlight these validation results, we have added explicit references to Fig. S2 in the method section. The corresponding method section now reads: "*... Additionally, the empirical coverage analysis (Fig. S2, top panels) provides a direct validation of Dis_UN's calibration, demonstrating that the 95th percentile predictions achieve their target coverage of approximately 95% across all traits. (Lines 201-204) "* Changed the title of Fig. S2: *"Validation of Dis_UN calibration and sensitivity analysis across quantile levels (τ = 0.75–0.99) for six plant traits: Leaf Mass per Area (g/m²) = LMA, Leaf Area Index (m²/m²) = LAI, Nitrogen content (mg/cm²) = N, Chlorophyll content (µg/cm²) = Chl, Equivalent Water Thickness (mg/cm²) = EWT, Carotenoid content (µg/cm²)= Car. Top panels: Empirical coverage (proportion of residuals falling within predicted bounds) with 68% Wilson confidence intervals, compared against*

		target quantile (1:1 line). At τ=0.95, coverage reaches ≈94-96% across all traits, validating model calibration. Bottom panels: Exceedance mean (average magnitude of errors exceeding predicted bounds), reflecting worst-case error severity. Smooth behavior up to τ=0.95 transitions to abrupt instability at τ≥0.97, particularly for EWT and Car, indicating entry into an unstable tail regime. Together, these results support τ=0.95 as a robust, data-driven choice that achieves high coverage while avoiding excessive conservatism and instability at extreme quantiles.
2	Further Clarify the Comparison with Ens_prob_UN in the Discussion. The results (Fig. 3) show that Ens_prob_UN (probabilistic deep ensemble) performs exceptionally well in calibration (ENCE), even better numerically than Dis_UN in some cases. It is recommended to state more explicitly in the discussion and conclusion the differentiating advantages of Dis_UN compared to this strong baseline: (a) Dis_UN provides a clearly defined "conservative error bound" (worst-case bound), which is fundamentally different from a symmetric interval. (b) In landscape-scale OOD detection (Figs. 6 & 7), Dis_UN demonstrates higher contrast and separability (i.e., higher JM distance) when distinguishing non-vegetation pixels (like water, clouds).	We thank the reviewer for this important suggestion to more explicitly articulate Dis_UN's advantages relative to Ens_prob_UN. We have added new paragraphs in Discussion Section 4.1 and 4.2.1 as well as a few lines in the conclusion and abstract that explicitly highlight the advantage of Dis_UN. The discussion now reads: ***"4.1 Local-Scale Uncertainty in OOD Vegetation data*** *Applying the ensemble and Monte Carlo methods with their default settings resulted in a critical underestimation of observed residuals (Fig. S6). After scaling variance-based uncertainties to approximate 95% confidence intervals (1.96 × σ), the ensemble approaches showed substantially improved alignment with observed residuals (Fig. 3), in contrast to their unscaled performance shown in Fig. S6. The improved performance of both ensemble methods for OOD vegetation samples indicates the average alignment under natural distribution shifts (Gustafsson et al., 2020), that is, when new data differ in source or environmental conditions but still represent the same underlying object class (vegetation). However, these methods do not automatically adapt to distributional changes if not validated with an independent set.* *This limitation highlights the need for post-hoc scaling and introduces practical challenges in operational settings: practitioners must determine appropriate scaling factors for each trait and application context, requiring either assumptions of normality or empirical calibration on held-out data, which may not be available for novel sensors or ecosystems. Recent studies have increasingly emphasized the need for robust recalibration strategies for variance-based uncertainty*

		methods, particularly under distributional shift (Liu et al., 2021; Ovadia et al., 2019; Palmer et al., 2022). While a comprehensive evaluation of such calibration strategies lies beyond the scope of this study, our results underscore that naïve application of variance-based methods without careful consideration of error distribution characteristics can lead to systematically biased uncertainty estimates. *However, the Monte Carlo approach continued to perform poorly (high ENCE values) even after scaling, indicating its failure to capture variability within samples from different vegetation types. The observed low alignment between predicted uncertainty and residuals suggests that the uncertainty estimates produced by these models do not fully represent the model's errors (Fig. 3). This underestimation, especially for higher predicted values, is a known limitation of Monte Carlo-based approaches (Hu et al., 2022; Klotz et al., 2022; Liu at al. 2021), which tend to be optimistic in their uncertainty estimates. (Lines 525-544)"* ***"4.2.1 Comparison with other methods*** *…Furthermore, the variation in spatial resolution reveals a critical operational advantage of Dis_UN. The distance-based method maintains robust OOD detection during sub-pixel variation. At 30m resolution (EnMAP), individual pixels frequently contain mixtures of vegetation and non-vegetation components, for example, urban pixels containing street trees, or forest edges mixing canopy and bare ground. Such mixed pixels exhibit intermediate spectral signatures that fall between pure scene components. For variance-based methods, particularly ensemble approaches, these mixed-signature pixels can produce moderate prediction variance that fails to clearly flag them as problematic, since ensemble members may converge on intermediate predictions with modest disagreement (Figs. 6, 7 and S8). This is evident in the low minimum KS values for ensemble methods at 30m (as low as 0.05), indicating poor separation for certain traits where mixed pixels dominate the scene. In contrast, Dis_UN's distance-based predictors explicitly measure dissimilarity relative to the training set, which consists predominantly of pure vegetation samples. Mixed pixels, even if spectrally intermediate, are recognized as dissimilar from the pure vegetation training manifold, resulting in elevated uncertainty predictions. This mechanism remains effective regardless of pixel purity, explaining Dis_UN's*

		consistent performance (mean KS: 0.648 at 30 m) and its particularly strong advantage over ensemble methods at coarser resolution (36% higher than Ens_prob_UN at 30 m vs. 69% higher at 1m). *At higher spatial resolution (NEON, 1 m), where less sub-pixel variation is present than in the EnMAP scene and component boundaries are sharper, the performance of variance-based methods improves relative to the medium-resolution case (EnMAP) (mean KS ≈ 0.33-0.47). However, despite this improvement, Dis_UN still achieved the highest separability (mean KS = 0.795), indicating a stronger and more consistent contrast between vegetated and non-vegetated components. Importantly, the improved ensemble performance at 1 m resolution remains resolution-dependent and cannot be assumed in operational medium-resolution satellite applications, which dominate current global monitoring systems (e.g., EnMAP, PRISMA at 30 m). (Lines 609-628)"* The Abstract and conclusions now read as follows: *"Compared to scaled variance-based methods, Dis_UN provides (1) a superior estimation of uncertainty in OOD scenarios, achieving 36% higher contrast (KS distances: 0.648 vs 0.475) between non-vegetation pixels, particularly under mixed-pixel conditions at medium resolution (30m); (2) uncertainty quantification without requiring normality or symmetry assumptions, accommodating asymmetric error patterns; (3) enhanced interpretability of uncertainty sources, as uncertainty is directly linked to sample dissimilarity from the training data; and (4) computational efficiency at inference (2.6-7.7× faster), requiring only a single forward pass compared to multiple passes for ensemble-based methods."*
3	Briefly Highlight Inference Efficiency in the Main Conclusion. The authors' discussion in the manuscript rightly notes that Dis_UN is faster at inference. This is a major practical advantage for large-scale remote sensing. This "train-once, predict-fast" feature is a key advantage over ensembles and deserves a brief mention in the main Conclusion (Section 5).	*We have changed the text accordingly. Now the* conclusion section reads: *"…Compared to scaled variance-based methods, Dis_UN demonstrates four key operational advantages: (1) superior estimation of uncertainty in OOD scenarios, particularly under mixed-pixel conditions at medium resolution (30m) common in operational satellite monitoring; (2) uncertainty quantification without requiring normality or symmetry assumptions, accommodating asymmetric error patterns; (3) enhanced interpretability of uncertainty sources, as*

		uncertainty is directly linked to sample dissimilarity from the training data; and (4) computational efficiency at inference (2.6-7.7× faster), critical for processing large-scale hyperspectral data. ."
4	Link LAI Saturation to Signed Residuals in the Outlook. The new discussion on LAI saturation is nice, as is the mention of analyzing signed residuals. It is recommended to link these two points in the Outlook (Section 4.5). The model's systematic underestimation of high LAI is a good example of a skewed, signed error. The outlook should suggest that future work could specifically model these biased error patterns, especially for saturation-prone traits.	*We have changed the outlook section accordingly. Now* Section 4.5 reads. *"Looking forward, further refinement of the distance-based method could involve testing on more diverse datasets and exploring hybrid approaches that combine it with complementary probabilistic techniques. In addition, while directional errors were not explicitly modeled in this study, analyzing signed residuals could help reveal trait- or vegetation-specific biases. The systematic underestimation of high LAI values due to spectral saturation (Fig. S17, Section 4.2.2) exemplifies such directional errors. We recognize this as a valuable avenue for future research and recommend that future developments in uncertainty modeling explore the use of signed residuals and the estimation of both lower and upper quantiles."*
5	The DI calculation uses the median of 50 nearest neighbors. Please add a brief justification for this choice of k, as it could influence the stability of the DI metric.	Thank you for the comments, we added a clarification in Section 2.1.2 "***2.1.2 Dissimilarity indices (predictors)*** *…We chose 50 neighbors as a compromise between preserving local similarity and avoiding excessive signal dilution, given the relatively small size of the training dataset (~7000 samples). Smaller values would lead to very fine-grained distance distributions that are overly sensitive to individual training outliers, while larger values progressively dilute the local dissimilarity signal by incorporating increasingly dissimilar samples. (Lines 225-228)"*
6	The metric QuRatio is defined in the methods but is not mentioned by name in the results. Please clarify if this corresponds to the 'range %' reported in the Figure 3 caption and Table S5 and ensure terminology is consistent.	Thank you for pointing this out. The metric referred to as "range %" in Figure 3 and Table S5 indeed corresponds to the QuRatio metric defined in the Methods section. We have revised the caption of Figure 3 and Table S5 and consistently used the term QuRatio throughout the manuscript to ensure clarity and consistency. Caption Fig.3 ***"Scatter plots comparing the predicted uncertainties (x-axis) from four methods— Distance-based (Dis_UN), probabilistic ensemble (Ens_prob_UN), deterministic ensemble (Ens_det_UN) and Monte Carlo dropout (MCdrop_UN) calibrated by a factor of 1.96 × standard***

		deviation—against observed residuals (y-axis) of the multi-trait models across six traits: Leaf Mass per Area (LMA), Nitrogen content, Chlorophyll content (Chl), Equivalent Water Thickness (EWT), Leaf Area Index (LAI), and Carotenoid content (Car). Each point represents a sample, colored by vegetation type. For each trait–method combination, the regression line (blue) is compared to the 1:1 line (black) to visualize alignment between predicted and observed errors. Model calibration is quantified by the Expected Normalized Calibration Error (ENCE), while the ratio of predicted to observed uncertainty ranges (QuRatio) indicates the coverage of residual variability (see also Table S4 and S5). Note that the default definition of uncertainty with the Ensemble and Drop out approaches did systematically underestimate the observed residuals (Fig. S6)."
7	The insight that urban pixels (EnMAP) have lower uncertainty than pure OOD (like water) because they are mixed with green spaces is important. This implies a challenge: the DI is diluted by spectral mixing. This should be briefly acknowledged in the discussion as a limitation for medium-resolution imagery.	We thank the reviewer for highlighting this important observation. However, we suggest reframing this not as a limitation but as a key operational advantage of Dis_UN for medium-resolution imagery. The reviewer is correct that urban pixels show lower uncertainty than pure non-vegetation pixels (e.g., water) due to spectral mixing with green spaces. However, this is the expected and desirable behavior: mixed pixels are partially vegetated and therefore should show intermediate uncertainty between pure vegetation and pure non-vegetation. The critical question is whether uncertainty methods can distinguish these gradations reliably. To clarify this we have substantially revised Discussion Section 4.2.1 (lines 609-628) to explicitly discuss this resolution-dependent behavior. "***4.2.1 Comparison with other methods*** *…Furthermore, the variation in spatial resolution reveals a critical operational advantage of Dis_UN. The distance-based method maintains robust OOD detection during sub-pixel variation. At 30m resolution (EnMAP), individual pixels frequently contain mixtures of vegetation and non-vegetation components, for example, urban pixels containing street trees, or forest edges mixing canopy and bare ground. Such mixed pixels exhibit intermediate spectral signatures that fall between pure scene components. For variance-based methods, particularly ensemble approaches, these mixed-signature pixels can produce moderate prediction variance that fails to clearly flag them as problematic,*

		since ensemble members may converge on intermediate predictions with modest disagreement (Figs. 6, 7 and S8). This is evident in the low minimum KS values for ensemble methods at 30m (as low as 0.05), indicating poor separation for certain traits where mixed pixels dominate the scene. In contrast, Dis_UN's distance-based predictors explicitly measure dissimilarity relative to the training set, which consists predominantly of pure vegetation samples. Mixed pixels, even if spectrally intermediate, are recognized as dissimilar from the pure vegetation training manifold, resulting in elevated uncertainty predictions. This mechanism remains effective regardless of pixel purity, explaining Dis_UN's consistent performance (mean KS: 0.648 at 30 m) and its particularly strong advantage over ensemble methods at coarser resolution (36% higher than Ens_prob_UN at 30 m vs. 69% higher at 1m). *At higher spatial resolution (NEON, 1 m), where less sub-pixel variation is present than in the EnMAP scene and component boundaries are sharper, the performance of variance-based methods improves relative to the medium-resolution case (EnMAP) (mean KS ≈ 0.33-0.47). However, despite this improvement, Dis_UN still achieved the highest separability (mean KS = 0.795), indicating a stronger and more consistent contrast between vegetated and non-vegetated components. Importantly, the improved ensemble performance at 1 m resolution remains resolution-dependent and cannot be assumed in operational medium-resolution satellite applications, which dominate current global monitoring systems (e.g., EnMAP, PRISMA at 30 m. (Lines 609-628)"*