Quantifying Mangrove Extent and Uncertainty: An Embedding-Driven Approach for Southeast Asia and Papua New Guinea

Vanna Teck[1*], Chanarun Saisaward[1], Weraphong Suaruang[1], M. Warizmi Wafiq[1], Andrea McMahon[1], Pimchanok Buapet[2], Daniel Marc G. dela Torre[1], Karis Tenneson[1], David Saah[1], and Ate Poortinga[1*]

[1]Spatial Informatics Group, LLC, 2529 Yolanda Ct., Pleasanton, CA 94566, USA
[2]International Blue Carbon Institute, 42A Boat Quay 049832 Singapore
[*]Address correspondence to: vteck@sig-gis.com; apoortinga@sig-gis.com

Response to Referee's Comments (RC2)

Dear Referee 2,

We sincerely appreciate your careful evaluation of our manuscript and your constructive and technically insightful feedback. Thank you for the time and effort you invested in reviewing our work. Your comments and suggestions have helped us substantially improve the clarity and methodological presentation of the manuscript, particularly regarding the contribution of AlphaEarth Foundation (AEF) embeddings, the distinction map-based area estimates (MBAE) and sample-based area estimates (SBAE), the motivation for the spatial similarity analysis, the sampling design, and the terminology used for reference sample units.

We have carefully considered and addressed each of your comments. Our detailed, point-by-point responses are provided below, and all corresponding changes in the revised manuscript have been highlighted.

Referee comments are presented in black text, and our responses are provided in red text.

General Comments

Referee Comment:

The manuscript presents AE embeddings with 64-dimensional features combined with gradient boosting to generate annual mangrove maps. While the authors confirm that this approach achieves the greatest classification stability, they do not provide evidence that it improves conventional S1/S2-based methods or existing global maps such as WorldCover in terms of mangrove delineation. Before introducing a complex, large dataset (64 layers) as a go-to method for mapping, its efficiency relative to simpler approaches needs to be demonstrated.

Response:

Thank you for this important comment. We agree that the present study did not directly compare AEF embeddings with conventional Sentinel-1/Sentinel-2 feature-based approaches or existing global products such as WorldCover. Therefore, no conclusion can be drawn from this study regarding whether AEF provides superior mangrove delineation accuracy or efficiency relative to these approaches. The objective of this study was instead to evaluate the temporal consistency, spatial structure, and applicability of AEF embeddings for regional mangrove monitoring across SEA and PNG. To address the reviewer's concern, we have revised the manuscript to clarify this scope and removed statements that could imply superiority or greater computational efficiency relative to conventional approaches. We also explicitly acknowledge the absence of direct benchmarking as a limitation and identify controlled comparisons with Sentinel-1/Sentinel-2 features and existing global products as an important direction for future research. The corresponding revisions have been made in the Discussion (Lines 496–502) and the Conclusion, where claims implying greater computational efficiency or superiority over traditional approaches have been removed.

Referee Comment:

The manuscript mentions integrating an embedding-based map with design-based area estimation (line 99). It also mentions correcting the map-based area estimation with design-based area estimation. It is not clear where the integration is performed or what it entails. The manuscript seems to suggest that the design-based area estimation corrects the map-based pixel counting and thereby integrates the two. In reality, these two are different approaches to estimating areas. The map-based approach is subject to mapping error and therefore biased. The design-based (sample-based) approach is unbiased as it uses statistical probability sampling. The Olofsson formulas for area estimation, in fact, only use the reference (sampled data) equation 6. Hence, this is a sample-based area estimation under design-based inference. So there is no integration of design-based area estimation and AE-embedding-based maps, other than that the generated maps are used for stratification to draw the reference sample. If one wants to integrate two estimates (map-based and sample-based), model-assisted area estimation is a way to increase the precision of the area estimation. These differences should be clearly presented in this manuscript and the "integration" needs to be clarified.

Response:

Thank you for this important clarification. We agree that our previous wording could imply that MBAE derived by pixel counting were integrated with or corrected by design-based area estimates. This was not the approach used in this study. The AEF-derived maps were used to define five strata for a stratified probability sampling design, and pixel counts within these strata were used to determine mapped stratum areas for sample allocation prior to reference validation. Following validation, the stratified probability sampling design was used for design-based inference and uncertainty quantification following Olofsson et al. (2014). Thus, the final annual mangrove area estimates were not obtained by directly correcting the MBAE. Given the importance of this comment, we have revised the manuscript throughout to consistently distinguish MBAE from SBAE, clarify the role of the embedding-based maps in stratified probability sampling design under design-based inference, and remove wording that could imply integration of the two approaches.

Referee Comment:

Spatial similarity is assessed using a distance-based similarity metric and Monte Carlo. It is not clear why more common Moran's I and kriging variances are not used. In particular, the latter can clearly show the spatial dependency of the variable with clear spatial structure and range. In addition, the motivation of spatial similarity assessment and its link with the AE variables are not clear. Spatial similarity is highly dependent on underlying landscape homogeneity and will vary naturally from place to place. What value does this bring besides an area being spatially homogeneous or heterogeneous? In particular, this assessment is not compared against Sentinel 2 or other satellite-based similarity assessments; therefore, the value of this analysis or the advantage of using AE is not clear. Also, how is the temporal variability calculated? Please clarify.

Response:

Thank you for this comment. We agree that the motivation for the spatial similarity analysis was not sufficiently clear. The purpose of this analysis was not simply to characterize spatial autocorrelation or landscape homogeneity, but to examine the geographic transferability of AEF embeddings. This is particularly relevant because embeddings are often used for few-shot classification: if sample units of the same class have similar representations regardless of geographic location, a small number of labeled sample units from one region may generalize well elsewhere. Conversely, strong geographic structure in the embedding space would indicate that geographically distributed training data remain important. We therefore assessed how geographic distance changes with nearest-neighbor rank in the 64-dimensional AEF embedding space. This directly evaluates whether sample units that are most similar in embedding space are also geographically clustered. Moran's I and variogram/kriging approaches characterize spatial autocorrelation of spatial variables and would provide complementary information, but they do not directly address this question of geographic transferability in a multidimensional embedding space. We agree that, without comparison against Sentinel-2 or other representations, this analysis cannot establish whether the observed geographic dependence is unique to AEF or represents an advantage over conventional satellite features. We have revised the manuscript accordingly and clarified that the analysis characterizes geographic structure and potential transferability of AEF embeddings rather than demonstrating superiority over Sentinel-2. Temporal consistency was assessed separately from the geographic similarity analysis using the interannual similarity of AEF representations from 2017–2024, and the Methods have been revised to clarify how this was evaluated.

Referee Comment:

The sample size calculation and allocation need further explanation; what was the sample size per region? What was the expected accuracy or standard error used for estimating and allocating sample size? Equation 7, what is (p_h)? What value did you use?

Response:

Thank you for this important comment. We agree that the sample-size calculation and allocation required further clarification. We have revised the Methods to explain the sample-size determination and allocation procedure and to define the terms used in the equations. A target standard error of 0.01 was specified for sample-size determination. Stratum-specific anticipated proportions (p_h) were specified prior to reference interpretation and ranged from 0.01 to 0.60 across strata and study regions, reflecting differences in anticipated within-stratum variability. We have also clarified that p_h represents the anticipated proportion used for sample allocation, whereas $\hat{p}_h$ represents the proportion estimated from the validated probability sample. The final numbers of probability sample units (n_h) allocated to each stratum by country and sampling region are now explicitly reported in Supplementary Table 17. The corresponding revisions have been made in Section 2.7 (Accuracy Assessment and Area Estimation; Lines 290–298).

Referee Comment:

A multi-year sample was collected; are they different (in terms of spatial location) for each year, or are they collocated? In other words, only once sample locations were selected, and the interpretation done for each year? Please clarify. If for the latter, change in mangrove area can be estimated, it would be informative to see the change area with uncertainty estimates.

Response:

Thank you for this comment. We clarify that independent sample locations were not selected for each year. A stratified probability sampling design was drawn from the five strata (S1–S5), and the same spatial sample units were retained across the 2017–2024 monitoring period. Each sample unit was interpreted using year-appropriate high-resolution reference imagery to determine mangrove/non-mangrove status, identify whether a change occurred, classify the change as gain or loss, and, where possible, identify the year of change. The annual reference labels were therefore derived from the same collocated sample units and used for year-specific accuracy assessment and area estimation. Regarding the reviewer's suggestion, the collocated multi-year reference design provides the information needed to characterize gain and loss through time. However, the present analysis reports annual mangrove extent and associated uncertainty and does not separately report design-based gain and loss area estimates with confidence intervals (CIs). We acknowledge design-based inference of gain and loss areas with associated uncertainty as a valuable extension for future work. The corresponding clarification has been added to Section 2.7 (Accuracy Assessment and Area Estimation; Lines 304–311), and the reference interpretation procedure is summarized in Figure 3 (page 16) of the revised manuscript.

Specific Comments

Referee Comment:

Please provide evidence or reference for this statement (Lines 52–55).

Response:

Thank you for this comment. We have added references to support this statement. Previous studies have demonstrated that MBAE derived directly from classified maps through pixel counting can be affected by classification error and may differ from SBAE obtained using a stratified probability sampling design and design-based inference (Olofsson et al. 2013; Kleinewillinghöfer et al. 2022; Tyukavina et al. 2025b). The corresponding references have now been added to the revised manuscript (Lines 57–62).

Referee Comment:

Figure 2: Are the forest samples the training sample or a stratified random sample used for accuracy and area estimation?

Response:

Thank you for this comment. The samples shown in the original Figure 2 (now Figure 4) are validated reference sample units used for model development, not the stratified probability sampling design used for accuracy assessment and design-based inference. Following validation, the 64-dimensional AEF embedding vectors (A00–A63) were extracted for each reference sample unit for each year from 2017 to 2024 and projected using UMAP. Reference sample units positioned closer together indicate greater similarity in their embedding representations, while overlap or clustering across years indicates relative temporal consistency. We have revised the caption of Figure 4 to clearly distinguish these model-development reference sample units from the independent stratified probability sample used for accuracy and area estimation (page 17).

Referee Comment:

Table 2: Please show this table as a figure with simple error ranges on a bar; this way, the temporal fluctuations can be clearly seen. In fact this can be directly compared with the map-based pixel counting.

Response:

Thank you for this helpful suggestion. We have replaced Table 2 (now Figure 10) with a multi-panel bar figure showing annual SBAE with 95% CIs for 2017–2024. To enable direct comparison with MBAE, we also present MBAE across probability thresholds of 0.30–0.70. The comparison shows that the map-based estimates generally exceed the SBAE, although the magnitude of the difference varies with the selected probability threshold. Increasing the probability threshold generally reduces the MBAE and brings it closer to the SBAE in several countries. This comparison highlights both the temporal variation in mangrove area estimates and the sensitivity of MBAE derived by pixel counting to the selected classification probability threshold. The corresponding threshold-specific MBAE values are provided in Section A.1 and A.4 of the Supplementary Material.

Referee Comment:

Table 3: Please provide the confidence ranges for the accuracy estimates. There is a systematic tendency of high omission errors in the classification. This needs to be clearly communicated and highlighted in the manuscript.

Response:

Thank you for this comment. We have revised Table 3 to include the 95% CIs for the accuracy estimates. We have also revised the manuscript to explicitly acknowledge the systematic tendency toward lower producer's accuracy (PA), indicating higher omission errors in the mangrove class. These omission errors indicate that some reference mangrove areas were classified as non-mangrove and may therefore contribute to underestimation of mangrove extent in the MBAE. These revisions are provided in Section 3.3.3 (Accuracy Assessment; Lines 400–401 and 403–404; page 24) of the revised manuscript.

Referee Comment:

Terminology: "samples" versus "sample units" Samples: multiple times, the word "samples" was used when referring to the sample size or sample locations. A sample means a set of sample units (see the Tyukavina et al. 2025 glossary and terminology explanation). I recommend using the correct term sample units or sample sites rather than samples, as I believe this study does not use multiple sets of samples.

Response:

Thank you for this clarification and for referring us to the terminology provided by Tyukavina et al. (2025). We agree and have revised the manuscript throughout to use "sample units" when referring to individual observations or locations and "sample" only when referring to the complete set of sample units. This terminology has also been updated in the relevant figures, tables, and captions for consistency.

References

Kleinewillinghöfer, Lukas, Pontus Olofsson, Edzer Pebesma, Hanna Meyer, Oliver Buck, Christian Haub, and Beate Eiselt (2022). "Unbiased Area Estimation Using Copernicus High Resolution Layers and Reference Data". In: *Remote Sensing* 14.19, p. 4903. DOI: `10.3390/rs14194903`.

Olofsson, Pontus, Giles M Foody, Martin Herold, Stephen V Stehman, Curtis E Woodcock, and Michael A Wulder (2014). "Good practices for estimating area and assessing accuracy of land change". In: *Remote Sensing of Environment* 148, pp. 42–57. DOI: `10.1016/j.rse.2014.02.015`.

Olofsson, Pontus, Giles M. Foody, Stephen V. Stehman, and Curtis E. Woodcock (2013). "Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation". In: *Remote Sensing of Environment* 129, pp. 122–131. DOI: `10.1016/j.rse.2012.10.031`.

Tyukavina, Alexandra, Stephen V. Stehman, Giles M. Foody, Sophie Bontemps, Linda See, Pontus Olofsson, et al. (2025a). *Land Cover and Change Map Accuracy Assessment and Area Estimation Good Practices Protocol*. Good Practices Protocol. Version 0.1. Land Remote Sensing Protocol.

Tyukavina, Alexandra, Stephen V. Stehman, Amy H. Pickens, Peter Potapov, and Matthew C. Hansen (2025b). "Practical global sampling methods for estimating area and map accuracy of land cover and change". In: *Remote Sensing of Environment* 324, p. 114714.