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 (RC1)

Dear Referee 1,

We sincerely appreciate your thorough review and the constructive, insightful feedback you provided on our manuscript. Thank you for the significant time and effort you invested in evaluating our research. Your valuable suggestions have been instrumental in improving both the clarity and overall quality of our work.

Each of your points has been carefully addressed, and we have provided comprehensive, point-by-point responses below. For clarity, all revisions and corresponding updates within the modified manuscript have been explicitly highlighted.

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

Referee Comment:

Line 85, citation.

Response:

Thank you for this comment. We agree that this statement requires supporting references. We have therefore added recent studies demonstrating the use of AlphaEarth Foundation (AEF) embeddings for large-scale and multi-temporal Earth observation applications. Chaiyana et al. (2026) applied AEF embeddings to mangrove intactness and degradation monitoring from 2017–2024, while Cai et al. (2026) used AEF embeddings to produce annual 10-m cropland maps across Southeast Asia. We have added these references in Lines 96–97.

Referee Comment:

Why Gradient Boost Tree?

Response:

Thank you very much for this constructive comment. We agree that the rationale for selecting the Gradient-Boosted Tree (GBT) classifier should be clarified. GBT was selected because it is a robust ensemble learning method capable of modeling complex nonlinear relationships in high-dimensional AEF embedding features. Recent studies have demonstrated the suitability of tree-based ensemble methods for embedding-based remote sensing classification. In particular, Benavides-Martínez et al. (2026) evaluated AEF embeddings using Random Forest, GBT, XGBoost, and LightGBM, demonstrating the applicability of these algorithms for capturing nonlinear relationships in embedding-based features and remote sensing classification. Based on these characteristics, we selected GBT as an appropriate classifier for generating annual mangrove probability maps from the 64-dimensional AEF embeddings. We have revised the Methods section to clarify this rationale (Section 2.6, Lines 223–227).

Referee Comment:

Why 95% training and 5% testing split? Does this have logic behind?

Response:

Thank you very much for this constructive comment. We agree that the rationale for the 95%/5% training-test split should be clarified. In our study, the split was implemented independently at the country level. The primary rationale was to maximize the number and geographic representation of reference sample units available for training each country-specific classifier while retaining a country-specific hold-out subset for model-development diagnostics. This strategy is particularly relevant to our systematic evaluation of embedding behavior across heterogeneous tropical coastal systems, where environmental conditions and labeled-data availability vary among countries. Brown et al. (2025) evaluated AEF embeddings across different labeled-data regimes, including low-shot scenarios, demonstrating the utility of AEF embeddings for downstream mapping under varying levels of training-data availability. Accordingly, our 95%/5% allocation prioritizes the use of available country-specific reference information for model fitting while preserving a hold-out subset for diagnostic evaluation during model development. We have revised Section 2.6 (Embedding-Based Classification) to clarify the rationale for the country-level 95%/5% training-test split (Lines 241-248).

Referee Comment:

How many samples of each country?

Response:

Thank you very much for this constructive comment. We agree that the number and distribution of reference sample units across countries should be clearly reported. To improve transparency and reproducibility, we have added a new table to the revised manuscript (Table 1) summarizing the number of reference sample units for each country included in the study. The table provides the country-level counts of reference sample units and their distribution across the reference classes, allowing readers to assess the representation of sample units used for model development across SEA and PNG. We have added this information to Section 2.3 (Lines 159–160) and Table 1 (page 6).

Referee Comment:

Figure 1, suggest revising as professional style if necessary.

Response:

Thank you very much for this valuable suggestion. We agree that improving the visual presentation of Figure 1 enhances the clarity and professionalism of the manuscript. Accordingly, Figure 1 (now Figure 2) has been redesigned with a cleaner layout, improved geographic context, and clearer representation of the study area across SEA and PNG (page 15).

Referee Comment:

Table 2. Checking entire table, Malaysia (2017 and 2018), why is this duplicated number?

Response:

Thank you very much for this careful observation. We rechecked the Malaysia results for 2017 and 2018 and confirm that the identical values are not a duplication error. The estimates were calculated independently for each year. The identical estimates are not a duplication error. Because the same probability sample units and stratum weights are used through time, annual SBAE changes only when the year-specific reference status of sampled units changes. No change affecting the weighted estimate was observed between 2017 and 2018 in Malaysia; consequently, the estimated mangrove area is identical for the two years.

Referee Comment:

Do you have any discussion about Table 2, why some years for a country suddenly drop, such as Cambodia (2019 and 2020)?

Response:

Thank you very much for this valuable comment. We investigated the apparent year-to-year fluctuations in the estimated mangrove area, including the decrease observed in Cambodia between 2019 and 2020. The observed difference is small relative to the uncertainty associated with the annual area estimates, and the corresponding 95% confidence intervals (CIs) overlap substantially. Because we did not directly estimate the variance of the paired year-to-year change, we do not interpret this apparent decrease as evidence of a statistically significant change in mangrove extent. Notably, the annual mangrove-area dataset provided by Zhang et al. (2026) also shows a decrease in mangrove area in Cambodia between 2019 and 2020, providing independent support for the direction of the observed temporal pattern. We have revised the Discussion accordingly and added supporting references to Olofsson et al. (2013) and Zhang et al. (2026). We have revised the corresponding discussion in Lines 461–467 to clarify this point.

Referee Comment:

Table 1, p stands for? Relevant S1–S5 considers individual country? Please simplify this.

Response:

Thank you very much for this constructive comment. We have revised Table 1 to improve clarity and simplify the terminology. Specifically, p_{t1} and p_{t2} are now defined as the predicted probability of mangrove occurrence for each pixel at the start of the analysis period (2017) and the end of the analysis period (2024), respectively. We also clarify that the S1–S5 stratification was applied independently to each country across SEA and PNG using the same probability thresholds ($L = 0.45$, $H = 0.55$). The table caption and corresponding description in Section 2.7 have been revised accordingly to make the stratification scheme easier to interpret (Lines 261–269).

Referee Comment:

Section 2.5, how GBT works? Please simplify in 1–2 sentences.

Response:

Thank you very much for this valuable suggestion. We have added a brief description of how the GBT algorithm works in Section 2.5 (now section 2.6). Specifically, GBT sequentially builds decision trees, with each new tree correcting errors from the preceding trees, and combines their predictions to produce the final classification output (Friedman 2001). This clarification has been added in Lines 227–229 of the revised manuscript.

Referee Comment:

Why was UMAP selected over other dimensionality reduction methods?

Response:

Thank you very much for this constructive comment. We acknowledge that several dimensionality reduction methods, such as t-SNE and UMAP, are available for visualizing high-dimensional embeddings. We selected UMAP because our objective was to qualitatively evaluate the separability and temporal consistency of embedding representations across spatial and temporal domains. UMAP provides a useful balance between preserving local neighborhood relationships and broader embedding structure while remaining computationally efficient for large, high-dimensional datasets, making it well suited for Earth observation embedding analysis. We have clarified this rationale in the revised manuscript, supported by Zhang et al. (2026) (Lines 186–190).

Referee Comment:

Adding a research diagram will be helpful for readers; this is highly suggested.

Response:

Thank you for this valuable suggestion. We agree that a research workflow diagram improves the clarity of the study. Accordingly, we have added a research framework diagram (Figure 1) that illustrates the overall methodology, including data preparation, embedding extraction, machine learning, validation, and analysis workflow (page 5).

Referee Comment:

In result section, the comparison with prior available mangrove products is highly recommended to emphasize how the embedding (EM) dataset is better to be considered for future study. Please also discuss.

Response:

Thank you very much for this valuable suggestion. We agree that comparison with existing mangrove products is important for evaluating the potential advantages of embedding-based area estimates. In the revised manuscript, we have expanded the Results and Discussion to contextualize the embedding-based area estimates relative to existing mangrove products, including Global Mangrove Watch (GMW) and the SERVIR Southeast Asia (SEA) mangrove dataset. Because these datasets contributed to the compilation of reference sample units in this study, they were not treated as fully independent datasets for quantitative benchmarking. We therefore do not make a direct performance comparison or claim that AEF is superior to these products. Instead, we highlight the consistent annual embedding representations provided by AEF from 2017–2024 and discuss their potential for multi-temporal mangrove monitoring. We further identify direct quantitative benchmarking against independent mangrove products, conventional Sentinel-1-, Sentinel-2-, and Landsat-based approaches, and satellite foundation-model frameworks such as Tessera and OlmoEarth as an important direction for future research. These revisions are provided in the Discussion (Lines 489–495 and 503–505).

Referee Comment:

Figure 3, color RGB contains value. If yes, this figure should have a legend.

Response:

Thank you for this helpful suggestion. We have revised Figure 3 (now Figure 5) by adding an RGB legend to clarify the color representation. The legend explains that the displayed RGB composite was generated from the transformed embedding features and is intended for visualization purposes only, thereby improving the interpretability of the figure for readers (page 18).

Referee Comment:

In the Discussion, this lacks support for the claim mentioned in the Abstract ("However, area estimates vary widely across studies, often because they rely on direct pixel counts without stratified sampling or design-based inference, making them sensitive to image and model biases that vary across space and time."). Please discuss why this approach is better than others.

Response:

Thank you very much for this insightful comment. We have expanded the Discussion to clarify the statistical advantages of using stratified probability sampling and design-based inference for area estimation compared with relying on map-based area estimates (MBAE) alone. Direct MBAE treat mapped class labels as the basis for area calculation and may therefore be biased by omission and commission errors. In contrast, our framework uses the embedding-based maps to define sampling strata, while the final mangrove area estimates are derived independently from the stratified probability sample using design-based inference, with sampling uncertainty quantified through 95% CIs following established good-practice recommendations (Olofsson et al. 2014). This provides statistically defensible, uncertainty-quantified area estimates and is particularly important for regional and multi-temporal monitoring where mapping errors may vary across space and time. We have incorporated this clarification into the revised Discussion (Lines 479–488).

Referee Comment:

This study can provide key message as recommendation to improve study. What limitation, and future improvement are?

Response:

Thank you very much for this valuable suggestion. We have revised the manuscript to more explicitly highlight the study limitations and future research directions. The main limitations include the lack of independent annual reference datasets covering the full 2017–2024 period and the evaluation of only one satellite foundation-model embedding framework. Future research should therefore benchmark AEF against conventional multi-sensor Earth observation approaches and emerging representation frameworks such as Tessera and OlmoEarth, supported by independent field observations and high-resolution reference data. We also identify opportunities to extend Earth observation embeddings beyond mangrove extent mapping to forest structural attributes such as tree canopy cover (TCC) and tree canopy height (TCH), aboveground biomass and carbon-stock estimation, seagrass extent and biomass mapping, and agricultural field-level crop classification based on embedding similarity to validated reference sample units. Potential operational applications, including EUDR-related commodity and plantation mapping and deforestation-risk assessment, are also identified. These limitations and future research directions have been incorporated into the revised manuscript (Lines 525–552).

Referee Comment:

In conclusion, "Compared with traditional approaches based on spectral indices…", did this study compare?

Response:

Thank you very much for this insightful comment. We agree that this study did not conduct a direct experimental comparison with traditional spectral index-based approaches. Therefore, we have revised the Conclusion to remove the comparative wording and avoid implying that a quantitative benchmark was performed. The revised text instead emphasizes the characteristics and potential of the embedding-based framework, particularly its use of pretrained, fixed-dimensional feature representations within a common embedding space to support temporally consistent annual mangrove mapping. Direct benchmarking against conventional spectral- and multi-sensor-based approaches is now identified as an important direction for future research. The corresponding revision is provided in Lines 559–561.

Referee Comment:

Suggest showing overview of mangrove extent at all countries.

Response:

Thank you very much for this valuable suggestion. We agree that a regional overview is helpful for visualizing the spatial distribution of mangroves across the study area. Accordingly, we have added a new multi-panel figure showing annual embedding-based mangrove forest probability maps across SEA and PNG from 2017 to 2024 (page 20).

References

Benavides-Martinez, I. F., J. Guthrie, J. E. Arias, Y. A. Garces-Gomez, A. I. Guzman-Alvis, C. V. Portilla-Cabrera, et al. (2026). "What on Earth is AlphaEarth? Hierarchical structure and functional interpretability for global land cover". In: *arXiv preprint arXiv:2603.16911*.

Brown, C. F., M. R. Kazmierski, V. J. Pasquarella, W. J. Rucklidge, M. Samsikova, C. Zhang, and P. Kohli (2025). *AlphaEarth Foundations: An Embedding Field Model for Accurate and Efficient Global Mapping from Sparse Label Data*. arXiv:2507.22291.

Cai, Y., B. Li, X. Liu, X. Jiang, Y. Zhu, S. Luo, et al. (2026). "Annual 10-m high-resolution cropland maps for Southeast Asia since 2019 using AlphaEarth embeddings". In: *Earth System Science Data Discussions* 2026, pp. 1–30.

Chaiyana, A., F. Sarvia, N. Nuthammachot, and J. Som-ard (2026). "A Scalable Potential of Alpha Earth and Multi-Sensor Datasets for Assessing Mangrove Intactness and Degradation using Deep and Machine Learning Algorithms". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Friedman, Jerome H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine". In: *The Annals of Statistics* 29.5, pp. 1189–1232. DOI: `10.1214/aos/1013203451`.

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`.

Zhang, G., S. Ding, and Z. Jin (2026). "A Comparative Study of UMAP and Other Dimensionality Reduction Methods". In: *arXiv preprint arXiv:2603.02275*.

Zhang, Z., N. J. Murray, X. P. Song, P. Bunting, T. A. Worthington, L. Fatoyinbo, et al. (2026). "Unexpected Expansion and Regrowth in Earth's Mangrove Forests over the Past Four Decades". In: *Science* 392.6802, pp. 1082–1087.