https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

Annual 10-m high-resolution cropland maps for Southeast Asia since 2019 using AlphaEarth embeddings

Yaotong Cai[1,2], Baoni Li[1], Xiaoye Liu[1], Xin Jiang[1,2], Yakun Zhu[1,3,4], Shuxin Luo[1,3,4], Yingzuo Qin[1], Shudi Xie[1,2,3], Jianhuai Ye[1], Huizhong Shen[1], Zhilin Guo[1], Xiaoping Liu[5], Zhenzhong Zeng[1,2]*

[1]School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China
[2]State Key Laboratory of Soil Pollution Control and Safety, Southern University of Science and Technology, Shenzhen, China
[3]School of the Environment and Sustainable Engineering, Eastern Institute of Technology, Ningbo 315200, China
[4]Zhejiang Key Laboratory of Industrial Intelligence and Digital Twin, Eastern Institute of Technology, Ningbo 315200, China
[5]School of Geography and Planning, Sun Yat-Sen University, Guangzhou, 510275, China

Correspondence to: Zhenzhong Zeng (zengzz@sustech.edu.cn)

Abstract. Southeast Asia (SEA) contributes substantially to tropical agriculture, but remains underserved by high-precision cropland data due to persistent cloud cover, fragmented farming, prevalent shifting cultivation, and complex phenology. Here, we developed a 10-m annual cropland dataset for SEA (SEA_Cropland10) covering 2019–2024 using a random forest model on Google Earth Engine. The model integrated 88,088 Sentinel-1 SAR scenes, 599,255 Sentinel-2 optical images, and Google AlphaEarth embeddings, and was trained using 37,192 visually interpreted samples. An independent accuracy assessment was performed using 1,200 samples stratified by land-cover change trajectories to validate both temporal dynamics and spatial extent. SEA_Cropland10 achieved an overall accuracy (OA) of 92.67% (±1.47%) for cropland dynamics, with annual static accuracies consistently exceeding 92.42% (±1.50%). The incorporation of AlphaEarth embeddings proved critical, improving model performance by 5.81%. Compared to existing global products (e.g., GLAD, WorldCereal), SEA_Cropland10 improved OA by 14.56%–19.97% and substantially enhanced the detection of sloping cropland (>5°) in mountainous regions of SEA, increasing producer's accuracy in global baselines from below 29.7%–32.9% to 89.0%–99.1%. Consequently, we identified three- to fourfold more sloping cropland area than GLAD and WorldCereal reported. Based on SEA_Cropland10, the estimated total cropland area in SEA shifted from 68.7 (±2.8) Mha in 2019 to 67.3 (±3.2) Mha in 2024, showing strong consistency with national statistics (r = 0.90–0.95). This dataset provides an important improvement for regional food security monitoring and carbon cycle modeling. The SEA-Cropland10 is publicly available at Zenodo: https://doi.org/10.5281/zenodo.17828801 (Cai and Zeng, 2026) and Google Earth Engine App: https://ee-caiyt33tc.projects.earthengine.app/view/seacrop10.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

1 Introduction

Southeast Asia (SEA) plays a pivotal role in the global food system, occupying approximately 3% of global agricultural land while producing about one-fifth of the world's rice, and serving as a major producer of commercial crops such as maize, cassava, and sugarcane (FAO, 2025; World Bank, 2025). This agricultural prominence, however, is embedded within one of the world's most biologically diverse and ecologically sensitive regions (Jung *et al.*, 2021; Struebig *et al.*, 2025). The region is currently undergoing rapid and often poorly regulated land-use change, characterized by the expansion of intensive monocultures into high-conservation-value forests and the encroachment of cultivation onto fragile highland slopes (Pravalprukskul *et al.*, 2023; Zeng *et al.*, 2018a, b). These processes have generated urgent environmental concerns, including accelerated soil erosion, altered hydrological disruption, and substantial carbon emissions due to deforestation and peatland loss (Feng *et al.*, 2022; Miettinen *et al.*, 2017; Valentin *et al.*, 2008). These intertwined agricultural and environmental pressures make spatially explicit, high-resolution, and temporally consistent characterization of cropland dynamics indispensable for understanding ongoing land-use change and supporting food security and sustainable land management in SEA.

The pressing need for agricultural monitoring has coincided with the open-data era, driving a proliferation of global land-cover and cropland datasets. Previously, researchers relied on coarse-to-moderate resolution LULC datasets, such as MODIS land cover (Friedl and Sulla-Menashe, 2022), and ESA Climate Change Initiative (CCI) land cover (ESA, 2017), which established essential baselines for global agricultural assessment. With the advancement of remote sensing, 30-m resolution products became widely adopted, including general LULC datasets such as GlobeLand30 (Chen *et al.*, 2015) and GLC_FCS30 (Zhang *et al.*, 2021), as well as specialized cropland products such as the NASA Global Cropland Extent Product (GCEP30) (Thenkabail *et al.*, 2021) and the GLAD Global Cropland (Potapov *et al.*, 2022). More recently, the Sentinel constellation has enabled the generation of 10-m resolution products, including general LULC products such as ESA WorldCover (Zanaga *et al.*, 2022), ESRI Land Cover (Karra *et al.*, 2021), and Google's Dynamic World (Brown *et al.*, 2022), as well as the specialized ESA WorldCereal system (Van Tricht *et al.*, 2023), which focuses specifically on global cropland and crop type mapping. Recent regional fine-resolution cropland mappings in SEA have primarily focused on specific crop types, such as paddy rice (Han *et al.*, 2021; Sun *et al.*, 2023), and therefore lack comprehensive cropland extent information, limiting their applicability for region-wide agricultural analyses. In contrast, global datasets exhibit substantial systematic errors and inconsistencies when applied to SEA, with spatial agreement reaching only 33% (Pérez-Hoyos *et al.*, 2017; Tubiello *et al.*, 2023; Waldner *et al.*, 2015).

A primary deficiency of global cropland maps lies in the scale mismatch and spectral complexity inherent to SEA's smallholder farming systems (Pérez-Hoyos *et al.*, 2017). Dominated by fragmented plots (often <0.5 ha) and intricate mixed cropping (Lesiv *et al.*, 2019; Waldner *et al.*, 2015), these landscapes exacerbate the mixed-pixel effect (Pérez-Hoyos *et al.*, 2017; Waldner *et al.*, 2015), a challenge further intensified by the region's vertical agricultural expansion (Zeng *et al.*, 2018b). In these rugged environments, topographic shadowing and the spectral similarity between active crops, shifting

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

cultivation cycles, and natural shrublands create substantial ambiguity for optical-based classifiers (Li *et al.*, 2014; Rufin *et al.*, 2022). Standard spectral features often fail to resolve these semantically distinct classes, causing global products prioritizing broad-scale consistency to systematically underestimate the agricultural footprint, particularly in sloping terrain (Miettinen *et al.*, 2014; Parker *et al.*, 2024). Furthermore, the intense monsoon climate leads to persistent cloud contamination, rendering optical imagery unreliable for continuous monitoring (Oliphant *et al.*, 2019). While multi-source fusion is a standard mitigation strategy, simple stacking of SAR and optical bands often fails to fully exploit the complementary information required to bridge frequent observation gaps (Pipia *et al.*, 2019). As a result, there is still no dedicated dataset that provides consistent, annual, and gap-free cropland maps for SEA, limiting the detection of abandonment and recultivation cycles and attribution of forest-to-cropland conversion timing.

Recent methodological advancements offer new avenues to address the limitations in feature representation under complex landscape constraints. Geospatial Foundation Models (GFMs), pretrained on massive Earth observations, have emerged as a promising approach to overcome the limitations in existing featurization methods (Brown *et al.*, 2025; Feng *et al.*, 2025; Li *et al.*, 2026; Zhu *et al.*, 2026). Representative frameworks, such as AlphaEarth, enable the extraction of high-level, self-supervised semantic features that capture complex contextual information and texture patterns often missed by traditional spectral indices or standard convolutional neural networks (Brown *et al.*, 2025). However, the effectiveness of GFMs for agricultural mapping remains under active debate. While internal evaluations suggest strong potential for general land cover mapping, recent independent benchmarking indicates that GFM embeddings may offer limited performance gains over optimized baselines in data-rich agricultural environments (Ma *et al.*, 2025). Consequently, the extent to which these models can generalize to data-sparse, heterogeneous tropical landscapes, where signal quality is the primary bottleneck, remains an open question requiring empirical validation.

In this study, we address the critical scarcity of reliable regional data by generating SEA_Cropland10, the first high-resolution (10-m), annual cropland dataset for SEA spanning from 2019 to 2024. Implemented within a cloud computing framework, our approach specifically targets the limitations of previous studies by: (1) integrating multi-source data (Sentinel-1, Sentinel-2) with AlphaEarth embeddings to enhance discrimination in fragmented and sloping landscapes; (2) employing a temporal consistency optimization strategy to reconstruct reliable annual trajectories despite cloud interference; and (3) providing rigorous validations against comprehensive reference samples, and comparative assessment with existing global products. This dataset aims to capture the previously overlooked dynamics of smallholder farming and upland agriculture, providing a refined baseline for agricultural and supporting environmental sustainability in SEA.

2 Study area and dataset

2.1 Study area

SEA, extending from approximately 10°S to 20°N latitude and 95°E to 140°E longitude, encompasses 11 countries (Thailand, Vietnam, Cambodia, Laos, Myanmar, Malaysia, Indonesia, the Philippines, Singapore, Timor-Leste, and Brunei)

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

with a total land area of approximately 4.5 million km^2 (Fig. 1a, b). The region features predominantly tropical monsoon and rainforest climates, characterized by high humidity, distinct wet and dry seasons, and abundant annual precipitation ranging from 1,500 to over 2,875 mm. Mean annual temperatures remain relatively stable between 25–28 °C, creating favorable conditions for year-round or multi-cropping agricultural systems.

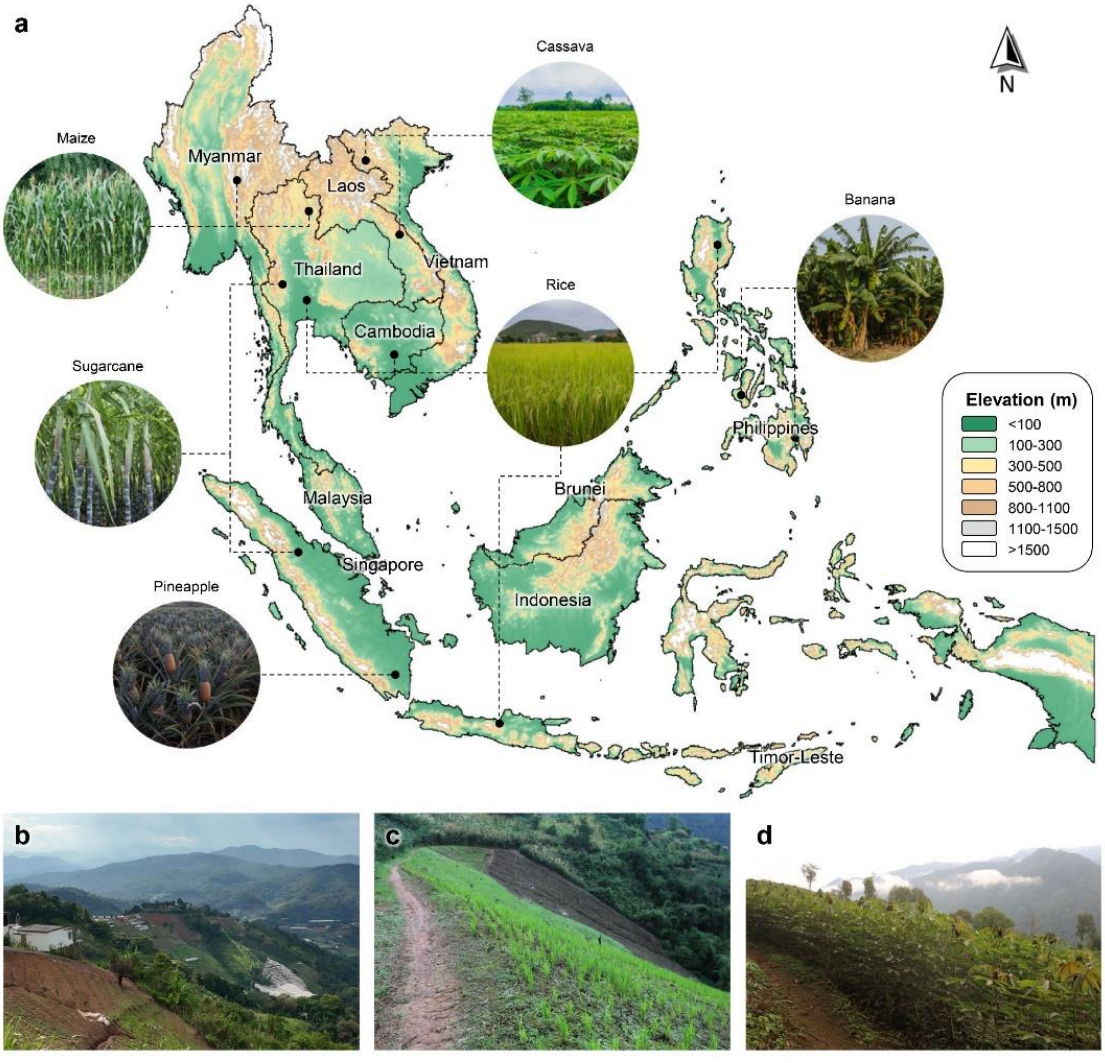


Figure 1: Overview of the Southeast Asia study area. (**a**) Regional topography overlaid with the locations of representative field photographs. (**b**)–(**d**) Typical landscapes of sloping agriculture captured in Nan Province, Thailand, in June 2025: (**b**) extensive hillside cropland, (**c**) upland rice cultivation on steep terrain, and (**d**) cassava cultivation on slopes.

As one of the world's most dynamic agricultural regions, SEA supports a population of over 680 million people and plays a pivotal role in global food supply chains. Unlike agricultural landscapes in other major production zones, SEA is defined by distinct topographic and structural complexities. While lowlands support intensive irrigation systems, vast highland environments sustain widespread sloping and rainfed agriculture (Fig. 1c-e), which is closely linked to deforestation, shifting

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

cultivation, and land degradation processes (Feng *et al.*, 2021; Zeng *et al.*, 2018a, b). Furthermore, agricultural landscapes across SEA are shaped by the predominance of smallholder farming, resulting in highly fragmented, fine-scale, and dynamic land-use mosaics. The combination of high crop diversity, small field sizes, persistent cloud cover, and intense monsoonal variability makes SEA one of the most challenging regions globally for accurate, high-resolution cropland mapping.

2.2 Sentinel-1 and Sentinel-2 imagery

We employed a total of 599,255 scenes from the Harmonized Sentinel-2 Level-2A surface reflectance product to ensure radiometric consistency between Sentinel-2A and 2B. To address the challenge of persistent cloud contamination in SEA, we utilized the advanced Cloud Score+ algorithm for rigorous cloud and shadow masking (Pasquarella *et al.*, 2023). Cloud-free pixels were subsequently aggregated into seasonal median mosaics to generate spatially continuous and temporally consistent reflectance imagery. Based on these seasonal composites, a suite of spectral indices was calculated to characterize vegetation vigor, water content, and soil background properties. These indices include NDVI, NBR, three Red-Edge based indices (NDRE1–3), NDBI, SAVI, MNDWI, and NDSI. The detailed definitions and calculation formulas are provided in Table S1.

Complementing the optical data, we utilized 88,088 scenes of 10-m resolution C-band Synthetic Aperture Radar (SAR) imagery from the Sentinel-1 mission (Interferometric Wide swath mode). We processed dual-polarization (VV and VH) Ground Range Detected (GRD) data. Standard preprocessing was applied within Google Earth Engine (GEE), including thermal noise removal, radiometric calibration, and terrain correction using the SRTM DEM to address geometric distortions in mountainous terrain (Mullissa *et al.*, 2021). The backscatter coefficients (σ^0) were converted to decibels (dB). To minimize speckle noise while preserving phenological stability, we generated monthly median composites for both VV and VH polarizations. Additionally, the monthly Radar Vegetation Index for Sentinel-1 (RVI4S1) was calculated to enhance sensitivity to crop structure and soil moisture variations (Bhogapurapu *et al.*, 2022).

To construct robust temporal features, we applied the Whittaker smoother (Eilers, 2003) to the time series of all extracted features. This step aims to mitigate residual noise, ranging from atmospheric artifacts in optical composites to speckle fluctuations in SAR, and reconstructed consistent phenological profiles prior to classification.

2.3 AlphaEarth foundation model embeddings

The AlphaEarth Satellite Embedding Dataset is a large-scale, high-resolution benchmark designed to advance intelligent remote sensing analysis and multimodal Earth observation research (Brown *et al.*, 2025). It integrates multi-source and multi-temporal satellite data, including optical, radar, LiDAR, topographic, and climatic modalities, and employs advanced deep feature extraction models to generate globally consistent embeddings at a 10 m × 10 m spatial resolution. These embeddings encapsulate rich semantic, spatial, and temporal representations of the Earth's surface, facilitating efficient transfer learning and cross-sensor harmonization. In this study, we utilized these pre-computed features, hereafter referred to as "AlphaEarth embeddings" (or simply "AlphaEarth"), to capture high-level landscape semantics.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

2.4 Topographic variables

Topographic variables were derived from the Copernicus Global Digital Elevation Model (GLO-30), which is based on interferometric synthetic aperture radar (InSAR) data acquired by the TanDEM-X mission. The GLO-30 dataset offers global coverage at 30 m resolution, with a reported absolute vertical accuracy of <4 m and a horizontal accuracy of <4 m. Elevation, slope, and aspect layers were computed via the GEE platform to characterize terrain-induced variations in crop distribution, particularly in complex mountainous environments.

2.5 Reference data collection

2.5.1 Samples for model development

To support supervised classification, we constructed a comprehensive set of training and validation samples representing 2024 land-surface conditions across SEA. We first generated 30,000 spatially random points within the study area. A rigorous visual interpretation protocol was developed to ensure label accuracy. For each sample point, we determined the land cover status by integrating multiple data sources: (1) Very-high-resolution imagery: Google Earth history imagery was used to identify field geometry, texture, and landscape context. (2) Time-series spectral analysis: We utilized GEE to inspect the annual temporal profiles of Sentinel-2 spectral indices (NDVI) for the year 2024. This allowed us to distinguish cropland from natural vegetation based on seasonal phenology and to identify active cultivation versus fallow periods.

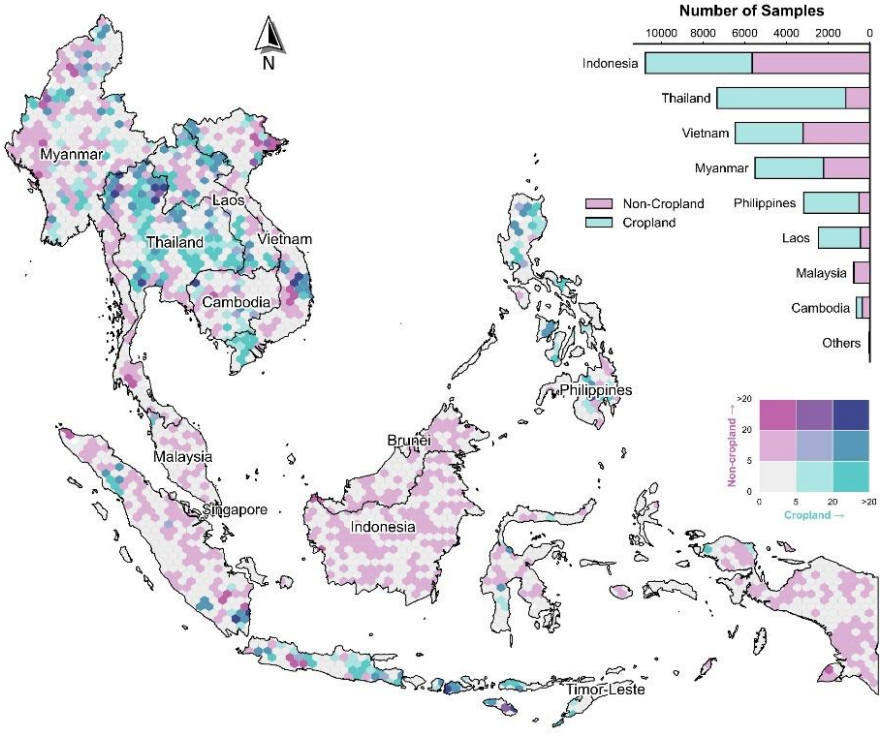


Figure 2: Spatial density and statistics of the samples for model development.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

To enhance the representation of complex agricultural landscapes, we further curated an additional set of challenging samples, including terraced fields, sloping cropland, and highly fragmented smallholder plots. These supplementary samples were specifically selected to reduce omission errors in heterogeneous terrain and mosaic agricultural systems. In total, the final reference dataset consists of 37,192 labeled samples (Fig. 2), comprising 22,821 cropland and 14,371 non-cropland points, all corresponding to the 2024 reference period.

2.5.2 Stratified random samples for map accuracy assessment

To ensure a rigorous evaluation of thematic accuracy and to derive unbiased area estimates of cropland dynamics, we adopted a design-based inference framework recommended by Olofsson *et al.* (2014). It is important to note that the validation dataset constructed in this section was completely independent of the training samples described in *Section 2.5.1*.

A stratified random sampling design was employed to collect validation reference data. To specifically evaluate model performance in complex terrain and capture rare change dynamics, we constructed the stratification layer based on the comparison between the start (2019) and end (2024) maps, integrated with slope data. This process resulted in five distinct strata:

1) *Flat stable cropland*: Areas classified as cropland in both 2019 and 2024 with a slope ≤ 5°.
2) *Sloping stable cropland*: Areas classified as cropland in both 2019 and 2024 with a slope > 5°. This stratum was explicitly separated to validate model robustness in mountainous agricultural systems.
3) *Stable non-cropland*: Areas consistently classified as non-cropland in both 2019 and 2024.
4) *Cropland gain*: Areas classified as non-cropland in 2019 but as cropland in 2024.
5) *Cropland loss*: Areas classified as cropland in 2019 but as non-cropland in 2024.

Guided by the sampling design principles of Cochran (1977), we targeted a standard error of 0.01 for overall accuracy. Although the theoretical minimum sample size required to meet this target was smaller, we established a robust total sample size of 1,200 units. This larger sample size was adopted to enable a disproportionate allocation strategy, ensuring sufficient representation for rare but critical strata such as sloping cropland and land-cover change categories. The samples were distributed as follows: 500 for stable non-cropland, 300 for flat stable cropland, 200 for sloping stable cropland, and 100 each for the cropland gain and loss strata (Fig. 3). This allocation guarantees that the confidence intervals for specific strata remain sufficiently narrow for reliable inference.

Reference labels were assigned using the same visual interpretation data sources (Google Earth and GEE time-series tools) described in *Section 2.5.1*. However, for this accuracy assessment, the interpretation scope was expanded to the entire study period. We examined the full temporal trajectory of each sample point from 2019 to 2024. By analyzing continuous NDVI time series and historical imagery, we identified the exact timing of land cover conversions. Consequently, we assigned a precise annual status (cropland or non-cropland) to each of the 1,200 sample units for every year, resulting in a comprehensive validation dataset comprising 7,200 annual reference observations. This intensive protocol supports both the verification of annual maps and the calculation of unbiased area estimates using stratified estimators.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

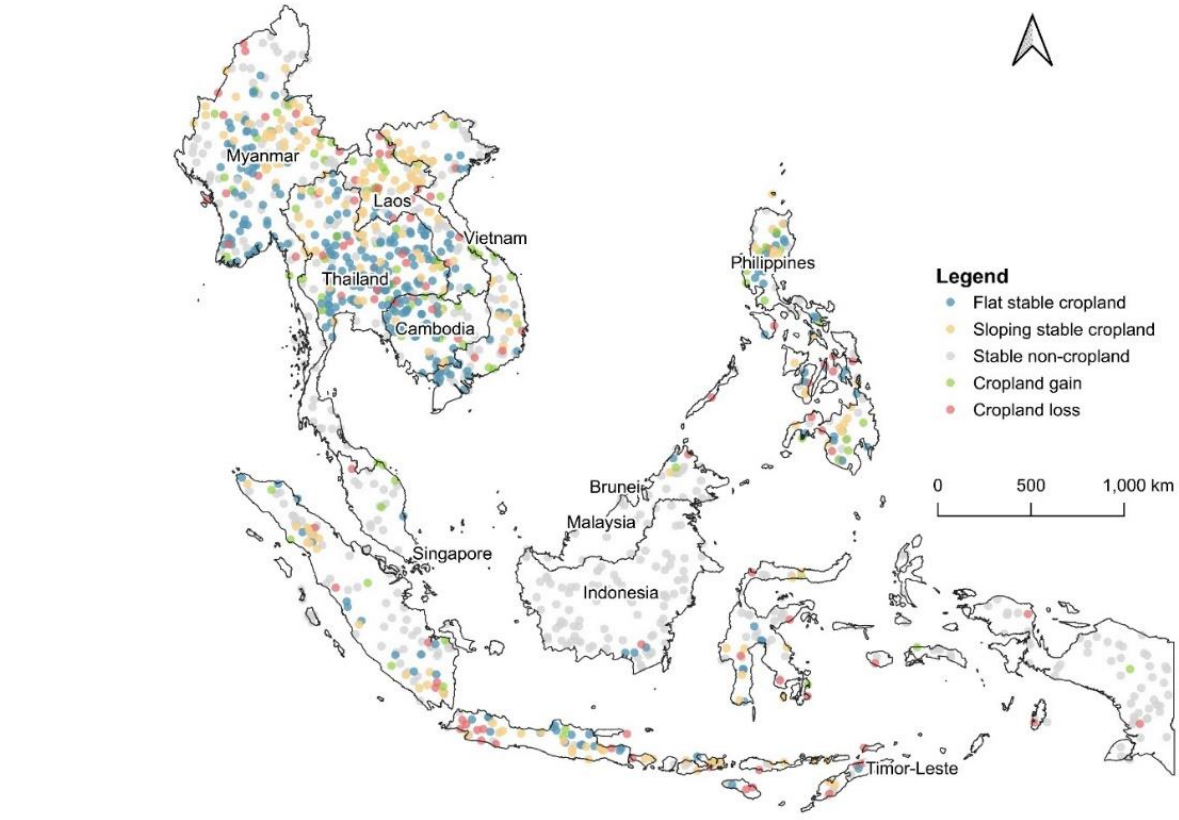


Figure 3: Spatial distribution of stratified random samples for accuracy assessment.

2.6 Official agricultural statistics

To assess the reliability of our remote sensing-derived area estimates at the national level, we utilized official agricultural statistics from the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT, 2023). Specifically, we extracted the annual cropland area statistics for Southeast Asian countries covering the period from 2019 to 2023. As the authoritative global standard for agricultural data, these records serve as an independent benchmark for evaluating the general magnitude and inter-annual trends of the derived map products across the region.

2.7 Cropland products for comparison

To evaluate the reliability and spatial consistency of our dataset, we selected three authoritative global cropland products as benchmarks. These datasets were chosen for their widespread community adoption and rigorous validation standards. While varying in spatial resolution and temporal epochs, they collectively represent the state-of-the-art in global agricultural monitoring. (1) The Global Cropland Extent Product (GCEP30), produced as part of the NASA Global Food Security-support Analysis Data (GFSAD) project (Thenkabail *et al.*, 2021). We utilized the sole available 2015 dataset, which maps global cropland extent at a spatial resolution of 30 m. (2) The WorldCereal global cropland map developed by the European

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

Space Agency (ESA) (Van Tricht *et al.*, 2023). Specifically, we utilized the temporary crops layer (v100), which provides a high-resolution (10 m) representation of active annual cropland for the year 2021. (3) The Global Cropland Extent dataset developed by the Global Land Analysis and Discovery (GLAD) laboratory at the University of Maryland (Potapov *et al.*, 2022). For this comparative analysis, we employed the latest available 2019 baseline product at 30 m spatial resolution.

3 Methods

In this study, we developed an automated and robust cropland mapping framework implemented on the GEE cloud computing platform to generate annual 10-m cropland maps for SEA from 2019 to 2024 (Fig. 4). The overall workflow integrates multi-source earth observation data, including Sentinel-1 SAR, Sentinel-2 optical imagery, and Google AlphaEarth satellite embeddings, to overcome challenges associated with persistent cloud cover and complex tropical phenology. The methodology is structured into four primary components: (1) data preprocessing and the generation of diverse feature sets; (2) collection of reference samples based on visual interpretation for model training; (3) annual classification using a Random Forest (RF) machine learning classifier followed by spatiotemporal consistency optimization; and (4) a rigorous accuracy assessment and area estimation based on a stratified random sampling design. The following sections detail each component of this framework.

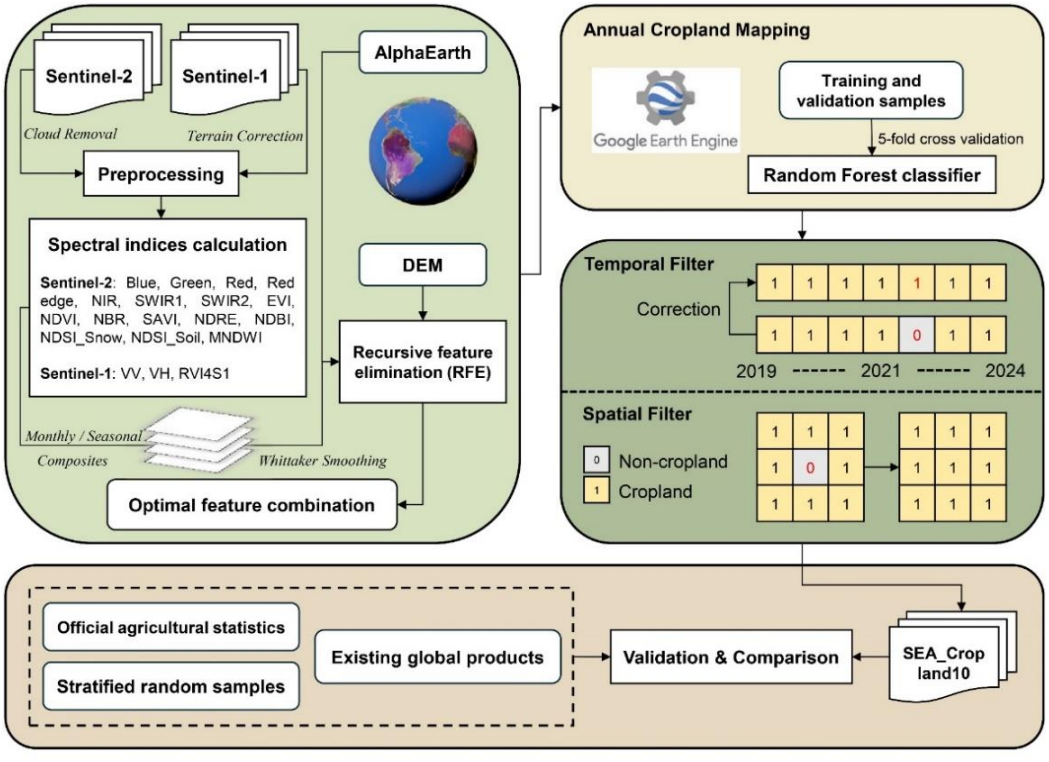


Figure 4: Overall technical flowchart of the Southeast Asia cropland mapping framework.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026


Open Access Earth System Science Data Discussions

3.1 Cropland definition

We defined cropland specifically as land covered by temporary herbaceous crops, which is adapted from the FAO Land Cover Classification System (LCCS) (Gregorio, 2005). This category encompasses cereals, vegetables, market gardens, and greenhouse-based cultivation. Consistent with FAO guidelines, semi-perennial crops such as sugarcane and cassava are included, as they are categorized as arable crops rather than woody perennials (Waldner *et al.*, 2015). To precisely characterize cropping intensity and shifting cultivation dynamics, our definition enforces two strict exclusions compared to broader arable land categories: (1) Exclusion of woody perennial plantations: Tree crops such as fruit orchards, rubber, and oil palm are excluded. These represent distinct land use systems with stable canopies that differ spectrally and structurally from seasonal cropland. (2) Exclusion of fallow land: Unlike standard agricultural statistics that often include temporary fallow as arable land, we classify all fallow periods, regardless of duration, as non-cropland. This physiognomy-based approach ensures that the dataset captures only active cropping events. By registering fallow periods as non-cropland, our annual map series effectively delineates the "cropland–fallow–cropland" cycles inherent to shifting agriculture, rather than masking them as stable land.

Finally, to characterize the relationship between agriculture and topography, cropland was further categorized based on terrain slope. According to the FAO slope classification standard (FAO, 2006), cropland located on terrain with a mean slope exceeding 5° was defined as sloping cropland, while that on terrain with a slope of 5° or less was classified as flat cropland.

3.2 Cropland classification approach

3.2.1 Feature optimization

We constructed a comprehensive initial feature stack consisting of 187 predictor variables, integrating multi-temporal Sentinel-1 backscatter, Sentinel-2 spectral indices and phenological metrics, AlphaEarth biophysical predictors, and topographic attributes derived from the DEM. To reduce feature dimensionality and mitigate the risk of overfitting, we applied a Recursive Feature Elimination (RFE) strategy guided by feature importance scores from a preliminary RF classifier (Cai *et al.*, 2024). Features were iteratively ranked and removed based on their importance contribution. At each iteration, a new model was trained and evaluated using five-fold cross-validation. The feature subset yielding the highest cross-validated accuracy was selected as the optimal combination for subsequent model development.

3.2.2 Model development and prediction

Building upon the optimal feature subset identified in Section 3.2.1, we employed a RF classifier within the GEE platform. This algorithm was selected for its computational scalability and robustness in handling high-dimensional remote sensing data (Breiman, 2017). The reference dataset collected in 2024 was utilized for model development and hyperparameter tuning via five-fold cross-validation.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

To ensure generalization across heterogeneous agricultural landscapes, hyperparameter tuning was conducted using a randomized grid search on the training partition. A total of 200 random configuration combinations were evaluated using five-fold cross-validation. The configuration yielding the highest validation accuracy was identified as the optimal model structure (Table S2). Subsequently, the final RF classifier was retrained using the complete training dataset to maximize data utilization. This fully trained model was then applied to the annual composites from 2019 to 2024 to generate pixel-wise cropland maps.

3.2.3 Model interpretability

To further enhance the transparency and interpretability of the RF model's internal decision-making mechanisms, we employed the SHapley Additive exPlanations (SHAP) method. Based on cooperative game theory, SHAP assigns each feature an importance value for a particular prediction, ensuring consistency and local accuracy (Lundberg and Lee, 2017). Specifically, we used the SHAP TreeExplainer, which is optimized for tree-based ensemble models, to calculate the marginal contribution of each feature across the dataset. This analysis allowed us to quantify the global contribution of different data sources and examine the directional impact of individual features on the probability of cropland detection.

3.2.4 Post-processing and spatiotemporal refinement

The optimized RF classifier was applied to the annual composites from 2019 to 2024 to generate initial pixel-wise binary classification maps. While the classification was robust, the raw outputs contained inherent noise attributable to cloud contamination, phenological variability, and spectral confusion. To enhance the dataset's reliability, we implemented a rigorous post-processing framework comprising temporal logical filtering and spatial morphological refinement.

Given the prevalence of shifting cultivation and short-term fallow systems in SEA (Li *et al.*, 2014), traditional temporal filters (e.g., moving window majority filters) risk over-smoothing genuine land use changes. To address this, we adopted a conservative trajectory-based anomaly correction strategy. We analyzed the land cover trajectory of each pixel over the entire six-year period (2019–2024) to identify and rectify illogical state transitions within stable cropland sequences. An anomaly was defined as a single year classified as non-cropland while the surrounding years (i.e., the remaining five years) were consistently classified as cropland. In such cases, the inconsistent year was reclassified as cropland. Crucially, this correction was strictly restricted to the intermediate years (2020–2023) to avoid introducing artifacts at the temporal boundaries. Furthermore, isolated cropland years were preserved rather than removed. This design choice explicitly maintains the integrity of shifting agriculture signatures, ensuring that genuine short-duration cropping events were not treated as noise.

Following temporal correction, spatial filtering was applied to remove "salt-and-pepper" noise and reconstruct coherent field structures. We utilized morphological operations based on connected component analysis to enforce a Minimum Mapping Unit (MMU) of approximately 0.1 ha (10 pixels). This process involved two steps: (1) Despeckling, where isolated cropland patches smaller than the MMU threshold were reclassified as non-cropland to reduce commission errors; and (2) Hole-filling,

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

where small non-cropland areas (<10 pixels) enclosed within large contiguous cropland patches were filled. Finally, the post-processed annual maps were exported as 8-bit integers with a consistent coordinate reference system.

3.3 Accuracy assessment and area estimation

To rigorously evaluate the reliability of the SEA_Cropland10 dataset and derive unbiased statistics, we employed a stratified inference framework based on the best practices guidelines (Olofsson *et al.*, 2014). This framework serves two purposes: (1) assessing the thematic accuracy of both the multi-year change dynamics and individual annual maps, and (2) estimating the unbiased cropland area with confidence intervals.

3.3.1 Accuracy metrics

A confusion matrix was constructed based on the interpreted reference samples. We calculated standard accuracy metrics, including overall accuracy (OA), producer's accuracy (PA, measuring omission error), and user's accuracy (UA, measuring commission error):

$$\mathrm{OA} = \frac{\sum_{i=1}^{n} x_{ii}}{N}, \tag{1}$$

$$\mathrm{PA}_i = \frac{x_{ii}}{\sum_{j=1}^{n} x_{ij}}, \tag{2}$$

$$\mathrm{UA}_i = \frac{x_{ii}}{\sum_{j=1}^{n} x_{ji}}, \tag{3}$$

where x_{ii} is the number of samples correctly classified for class i; x_{ij} is the number of samples of class i classified as class j; n is the total number of classes; N is the total number of validation samples; PA_i reflects the probability that a reference sample of class $i$is correctly mapped, indicating omission error; UA_i reflects the probability that a sample classified as class $i$actually belongs to that class, indicating commission error; OA represents the proportion of correctly classified samples across all classes.

We applied these metrics in two dimensions. (1) Dynamic Assessment: Evaluating the accuracy of the stable/change strata (e.g., stable cropland, gain, loss) over the 2019–2024 period. (2) Annual Assessment: Evaluating the static accuracy for each individual year by comparing the annual map classification against the reference label for that specific year.

3.3.2 Unbiased area estimation

Simply aggregating pixel counts from classified maps often results in biased area estimates due to classification errors. To address this, we estimated the area of cropland and its changes using the unbiased stratified estimator (Olofsson *et al.*, 2014). The estimated area of class j ($\hat{A}_j$) was calculated as:

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

$$\hat{A}_j = A_{tot} \sum_{i=1}^{k} W_i \frac{n_{ij}}{n_i}, \tag{4}$$

where A_{tot} is the total area of the study region; k is the total number of strata; W_i is the area proportion of map stratum i (derived from the classification map); n_i is the total number of samples allocated to stratum i; and n_{ij} is the number of samples in stratum i that were identified as class j in the reference data.

A key advantage of our sampling design is the ability to reconstruct unbiased annual trajectories using the same stratified framework. For estimating the cropland area for any individual year t (from 2019 to 2024), the stratification weights (W_i) remained constant, as they are defined by the stable/change strata from the aggregate map. However, the reference proportion (n_{ij}/n_i) varies annually, determined by the interpreted status of the samples in year t. This approach ensures that the area estimates are consistent and independent of the map's temporal artifacts (Potapov *et al.*, 2022).

3.3.3 Uncertainty analysis

To quantify the reliability of the derived results, we calculated the standard errors for both the area estimates and the accuracy metrics, constructing 95% confidence intervals for each indicator. For the area estimates, the standard error $S(\hat{A}_j)$ for class j was calculated as:

$$S(\hat{A}_j) = A_{tot} \sqrt{\sum_{i=1}^{k} W_i^2 \frac{\frac{n_{ij}}{n_i}(1 - \frac{n_{ij}}{n_i})}{n_i - 1}}, \tag{5}$$

Similarly, the standard errors for the accuracy metrics were computed based on the variance estimators for stratified random sampling derived by Olofsson *et al.* (2014). The 95% confidence intervals were then constructed as *Estimate* ± 1.96 × *SE*.

4 Results and discussions

4.1 Model performance and feature contribution

The quantitative assessment of model performance, demonstrates the robust predictive capability for cropland mapping in SEA (Table 1). The model integrating AlphaEarth achieved an OA of 95.79%, representing a significant improvement of 5.81% over the baseline model (90.53%) that relied solely on conventional satellite data. This performance gain is evident across all class-specific metrics. Notably, the PA for cropland increased from 89.82% to 95.33%, indicating a marked reduction in omission errors. Furthermore, the UA for non-cropland improved substantially from 82.83% to 92.38%, suggesting that the semantic features from AlphaEarth effectively suppress false positives in complex landscapes.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Open Access Earth System Science Data Discussions

Table 1. Model performance with (w/) and without (w/o) AlphaEarth satellite embedding.

Metrics	w/ AlphaEarth	w/o AlphaEarth
PA (cropland)	95.33%±0.31%	89.82%±0.25%
UA (cropland)	97.94%±0.23%	95.38%±0.30%
PA (non-cropland)	96.58%±0.37%	91.86%±0.49%
UA (non-cropland)	92.38%±0.53%	82.83%±0.45%
OA	95.79%±0.16%	90.53%±0.24%

To interpret the drivers behind this performance improvement, we analyzed the feature selection process and relative contributions (Fig. 5). First, the RFE process identified an optimal subset of 83 features from the initial pool of 187 to maximize model efficiency (Fig. 5a). Among these, 59 out of the 64 input AlphaEarth features were retained, corresponding to a high retention rate of 92%. In comparison, only 19 Sentinel-2 optical features, 4 Sentinel-1 SAR features, and 1 terrain feature were selected. This preference suggests that the classifier finds the high-dimensional representations from

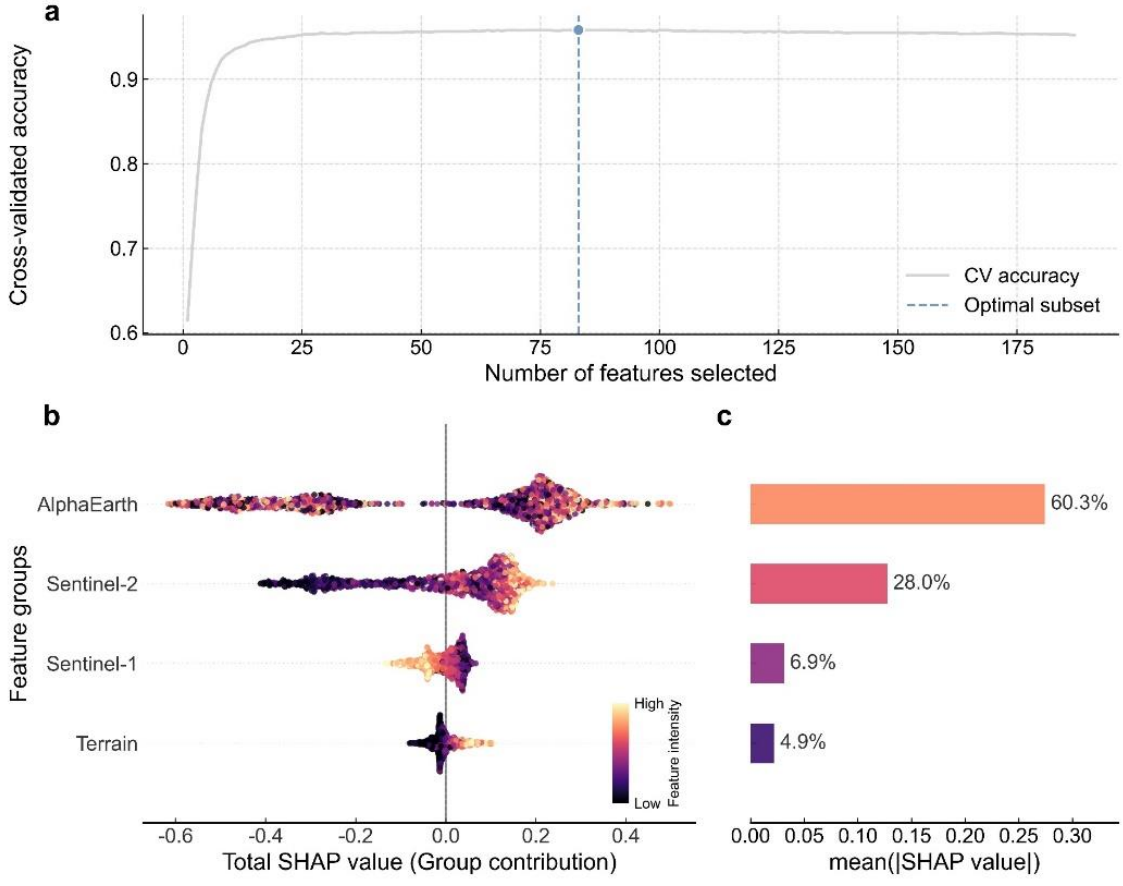


Figure 5: Feature selection results and model interpretability analysis. (**a**) RFE accuracy curve for optimal feature number determination. (**b**) Impact of aggregated feature groups on cropland classification. (**c**) Summary of feature importance and contribution rates by data source.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

AlphaEarth highly informative. This finding is corroborated by the SHAP-based importance analysis (Fig. 5c). AlphaEarth is the dominant driver of the classification, contributing 60.3% to the total feature importance. This is more than double the contribution of Sentinel-2 (28.0%), while Sentinel-1 and Terrain features played supplementary roles (6.9% and 4.9%, respectively). Finally, the SHAP beeswarm plot (Fig. 5b) reveals distinct behavioral patterns among data sources. The AlphaEarth group exhibits the widest spread, predominantly concentrated on the positive side (SHAP > 0), confirming its strong contribution to cropland identification. However, unlike physical features, AlphaEarth displays a complex, non-linear relationship (mixed colors), suggesting that it captures high-level semantic patterns rather than simple linear trends, which makes the direct physical interpretation of these features inherently challenging. In contrast, traditional features show clear linear correlations: Sentinel-2 and Terrain features exhibit a positive correlation, whereas Sentinel-1 features show a negative correlation.

4.2 Accuracy assessment based on stratified random sampling

The quantitative accuracy assessment of the SEA_Cropland10 product was conducted using validation samples stratified by both year and topography (Table 2). The product exhibited robust performance and high temporal consistency throughout the study period (2019–2024), with OA consistently exceeding 92% and peaking at 95.17% in both 2019 and 2024. Unlike many global land cover products that suffer from high omission errors in hilly terrains, SEA_Cropland10 maintained exceptional sensitivity in sloping regions. The PA for sloping cropland ranged from 85.00% (2022) to as high as 99.06% (2019), demonstrating that the majority of hillside farming activities were successfully captured. In several years, the detection rate for sloping cropland was comparable to, or even marginally higher than, that of flat cropland.

In terms of commission errors, the UA for sloping cropland remained stable between 85.71% and 88.46%. While slightly lower than the UA observed in flat regions (>94%), this is an acceptable trade-off for the significantly improved detection rate (PA) in complex terrains. For non-cropland classes, both flat and sloping regions showed high reliability, with UA and PA values generally exceeding 92%, indicating effective separation between cropland and other natural vegetation types such as forests or shrublands.

To further evaluate the product's capability in characterizing specific transition types, we calculated the accuracy metrics for five distinct strata: flat stable cropland, sloping stable cropland, stable non-cropland, cropland gain, and cropland loss (Table 3). Under this assessment framework, the product achieved a remarkable OA of 92.67% (±1.47%). The stable cropland classes exhibited exceptional reliability, with PA reaching 93.75% for flat regions and peaking at 98.19% for sloping regions. This reinforces the finding that the map is highly effective in identifying established agricultural zones across diverse topographies without significant omission errors. Regarding the dynamic classes (Cropland Gain and Loss), the product maintained high sensitivity, with PA values exceeding 92.50% for both categories. This indicates that the vast majority of cropland expansion and abandonment events were successfully captured. While the UA for these change classes (74.00%–79.00%) was slightly lower than that of stable classes, this is a typical characteristic of change detection tasks, attributable to the challenges in resolving exact change timing and mixed pixels at class boundaries. Nevertheless, the stratified assessment

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

confirms that SEA_Cropland10 provides a reliable dataset for monitoring agricultural dynamics in SEA, particularly effectively addressing the high cropland omission issue common in mountainous areas.

Table 2. Annual accuracy metrics of SEA_Cropland10 stratified by topographic gradients from 2019 to 2024.

Year	Metrics	Flat		Sloping		OA (%)
		Cropland	Non-cropland	Cropland	Non-cropland	
2019	PA (%)	96.65±1.86	94.49±2.81	99.06±1.30	92.00±2.75	95.17±1.21
	UA (%)	96.11±2.00	95.24±2.63	87.55±4.17	99.42±0.80	
2020	PA (%)	96.94±1.78	94.07±2.91	95.43±2.92	92.84±2.56	94.75±1.26
	UA (%)	95.87±2.05	95.58±2.55	87.04±4.48	97.58±1.56	
2021	PA (%)	94.44±2.43	93.70±2.90	89.01±4.54	93.35±2.42	93.08±1.44
	UA (%)	95.00±2.32	93.01±3.03	85.71±4.99	94.99±2.14	
2022	PA (%)	92.49±2.83	93.91±2.81	85.00±5.22	94.61±2.19	92.42±1.50
	UA (%)	94.77±2.42	91.29±3.26	87.43±4.91	93.46±2.38	
2023	PA (%)	93.51±2.62	93.77±2.87	91.62±3.93	93.70±2.39	93.33±1.41
	UA (%)	94.91±2.36	92.09±3.17	87.50±4.58	95.88±1.98	
2024	PA (%)	97.20±1.71	92.55±3.22	99.04±1.32	92.88±2.59	95.17±1.21
	UA (%)	94.81±2.27	95.93±2.47	88.46±4.09	99.44±0.78	

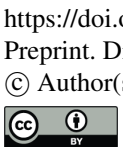
Table 3. Accuracy assessment of SEA_Cropland10 characterized by topography and change dynamics (2019–2024).

Class	PA (%)	UA (%)
Flat stable cropland	93.75±2.65	96.33±2.13
Sloping stable cropland	98.19±2.03	87.00±4.66
Stable non-cropland	90.35±2.47	99.20±0.78
Cropland gain	92.50±5.77	74.00±8.60
Cropland loss	92.94±5.45	79.00±7.98
Overall Accuracy (OA): 92.67%±1.47%		

4.3 Overviews of the SEA_Cropland10 dataset

Based on the SEA_Cropland10 dataset, the total mapped cropland area in SEA for 2024 is 67.7 Mha. The spatial distribution exhibits significant geographical heterogeneity (Fig. 6), characterized by a dominant concentration in the mainland corridors. The mainland SEA accounts for the majority of the region's agricultural footprint, contributing 73.3% of the total cropland area. Within this sub-region, Thailand holds the largest share (25.2%), followed closely by Myanmar (21.8%). Cambodia, Vietnam, and Laos account for 10.0%, 8.8%, and 6.8%, respectively, forming a continuous agricultural belt across the

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026



Indochinese Peninsula, while Malaysia contributes a smaller fraction (0.7%). In contrast, Maritime SEA comprises 26.7% of the total cropland extent. This distribution is primarily driven by Indonesia (16.7%) and the Philippines (9.7%), with Timor-Leste accounting for the remaining 0.3%. This distinct spatial pattern highlights the extensive, contiguous agricultural plains in the mainland deltas compared to the more fragmented, island-based agricultural systems in the maritime region.

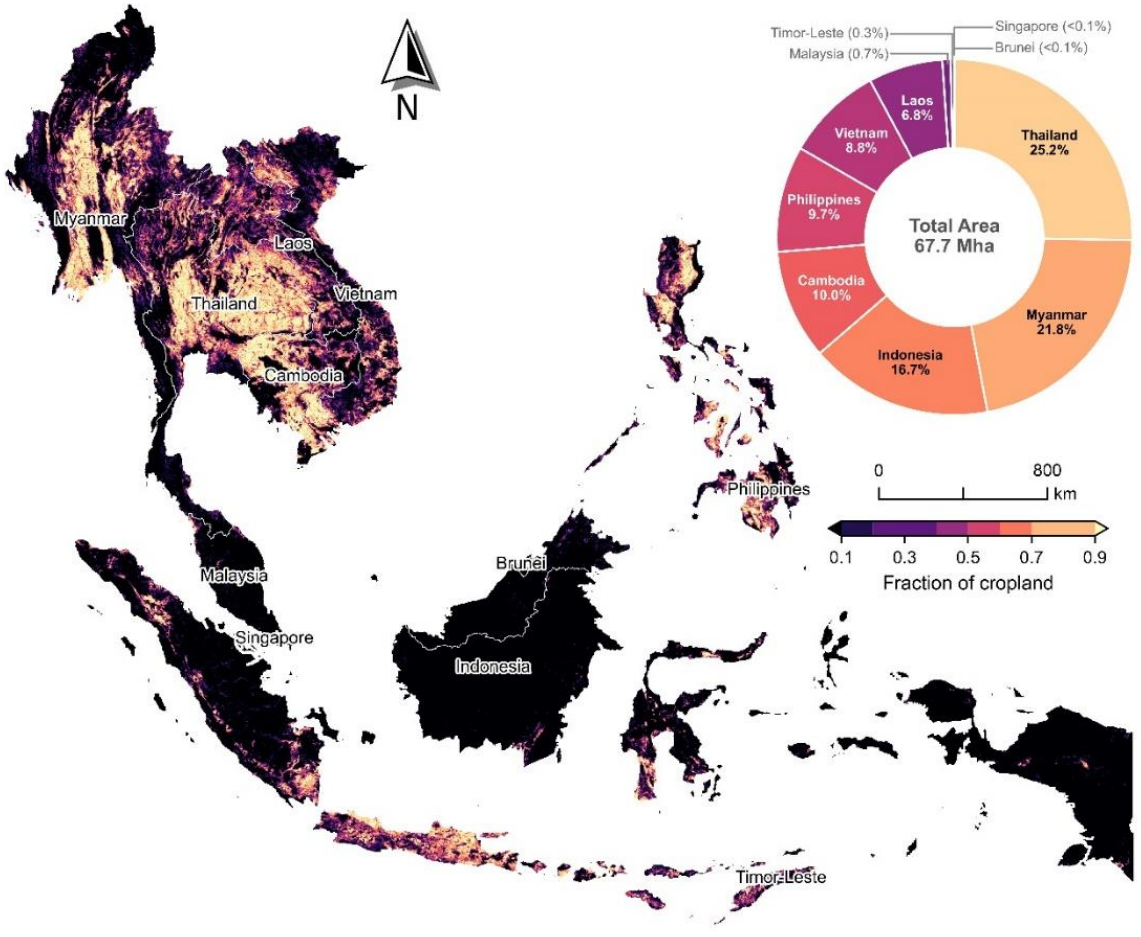
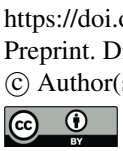


Figure 6: Spatial pattern of cropland in Southeast Asia in 2024. The main map displays the fractional cropland cover at a 2-km grid scale, aggregated from the 10-m binary classification map. The chart illustrates the proportional contribution of cropland area by country.

The spatiotemporal fidelity of the SEA_Cropland10 dataset is illustrated by mapping the temporal onset of cropping activities (Fig. 7a), a metric further substantiated by the corresponding Sentinel-2 time series (Fig. 7b–g). The dataset effectively delineates distinct land-use trajectories within complex landscapes. In mountainous terrains (Fig. 7a–e), the detection maps exhibit high spatiotemporal heterogeneity, forming a mosaic of asynchronous detection years. This pattern captures the rotational nature of shifting cultivation, where the dataset resolves agricultural dynamics down to the field scale, accurately distinguishing active cultivation phases from fallow periods. Conversely, Scene 6 captures the contiguous expansion of permanent agriculture, marked by the synchronized conversion of natural vegetation into cropland (Fig. 7g).

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026



Open Access Earth System Science Data Discussions

These distinct spatiotemporal signatures demonstrate the dataset's capacity to discriminate between the intermittent cycles of shifting agriculture and the systematic progression of permanent agricultural expansion.

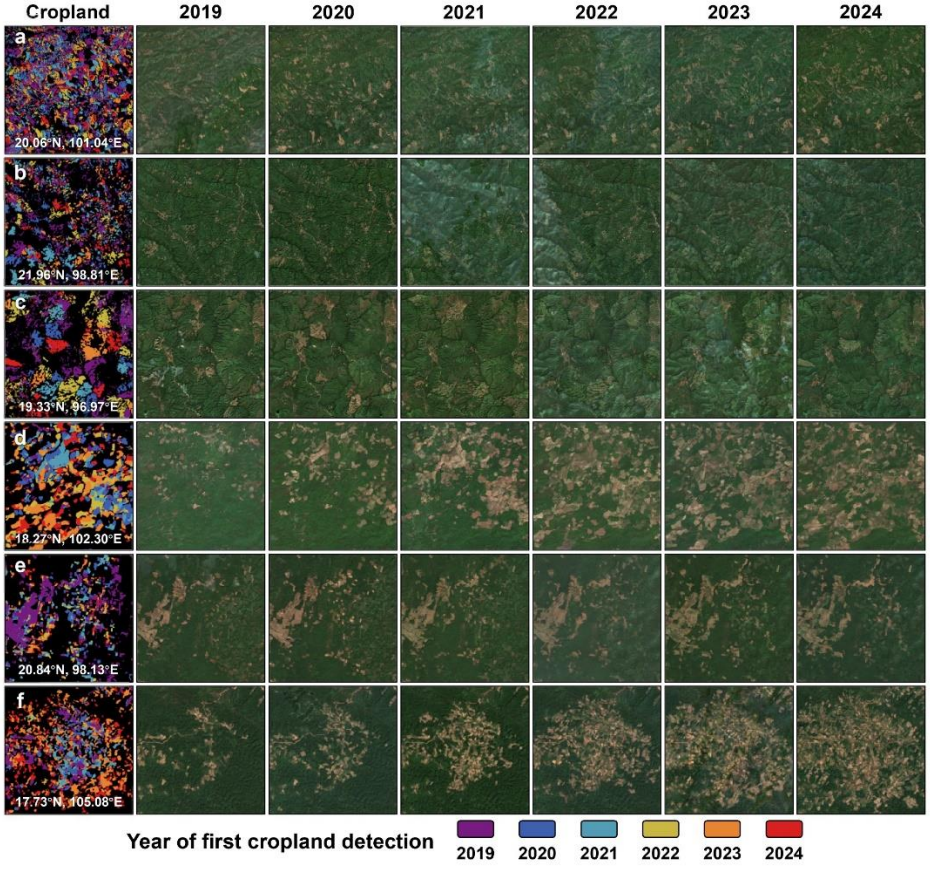


Figure 7: Examples of the first cropland detection year identified by SEA_Cropland10. Rows (**a**)–(**f**) represent different agricultural landscape scenes. The first column displays the spatial distribution of the first detection year, while the subsequent columns show the corresponding annual Sentinel-2 true-color composites from 2019 to 2024, providing visual reference for the detected land-cover changes.

The SEA_Cropland10 dataset demonstrates high quantitative reliability, with map-based area estimates aligning closely with unbiased sample-based inference (Table 4). Regionally, cropland area exhibited a distinct fluctuation rather than a monotonic trend, declining from a baseline of 69.4 Mha (2019) to a minimum of 59.4 Mha (2022) before recovering to 67.7 Mha (2024). This trajectory reflects substantial gross transitions characteristic of shifting cultivation systems: unbiased estimates indicate that cropland loss (10.8±1.2 Mha) marginally exceeded cropland gain (9.3±1.8 Mha) over the study period (Table 5), resulting in a slight net decline. Notably, our regional estimates surpass FAO statistics (~64 Mha), a discrepancy primarily attributable to the enhanced detection of sloping stable cropland (11.4±1.8 Mha) often omitted in traditional censuses.

Temporal trajectories varied substantially at the national level (Table 4). Laos recorded the most significant relative expansion, showing a continuous net increase from 3.4 Mha to 4.6 Mha. In contrast, Indonesia exhibited a consistent net

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

decline, decreasing from 13.4 Mha to 11.3 Mha. Meanwhile, major agricultural producers such as Thailand and Myanmar mirrored the regional "V-shaped" pattern, experiencing a temporary contraction in 2022 followed by a robust recovery to levels comparable to their 2019 baselines.

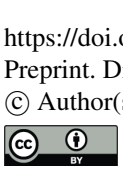
Table 4. Map-based annual cropland area estimates by country and regional comparison with sample-based and FAOSTAT data (2019–2024).

Region / Country	2019	2020	2021	2022	2023	2024
			Map-based Area (Mha)			
Brunei	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Cambodia	6.5	6.8	6.5	5.6	6.5	6.8
Indonesia	13.4	11.0	10.5	10.1	11.9	11.3
Laos	3.4	3.6	3.4	2.8	4.0	4.6
Malaysia	0.6	0.6	0.4	0.4	0.5	0.5
Myanmar	14.6	14.7	15.1	13.5	14.2	14.7
Philippines	6.7	6.9	5.8	5.7	5.8	6.5
Singapore	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Thailand	17.5	18.3	17.1	16.0	16.8	17.1
Timor-Leste	0.3	0.3	0.3	0.2	0.3	0.2
Vietnam	6.4	6.3	6.1	5.1	5.7	6.0
SEA	69.4	68.4	65.1	59.4	65.6	67.7
Sample-based Est. (SEA)	68.7±2.8	66.6±3.1	64.0±3.7	62.4±3.8	63.4±3.5	67.3±3.2
FAOSTAT (SEA)	64.1	64.2	63.6	63.3	63.2	—

Table 5. Map-based and sample-based annual cropland change estimates from 2019 to 2024.

Stratification class	Mapped area (Mha)	Est. area (Mha)
Flat stable cropland	44.6	46.5±2.6
Sloping stable cropland	11.9	11.4±1.8
Stable non-cropland	365.8	368.3±3.2
Cropland gain	11.2	9.3±1.8
Cropland loss	12.9	10.8±1.2

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026



Open Access Earth System Science Data Discussions

4.4 Comparison with existing datasets

At the national level, the estimated cropland areas derived from SEA_Cropland10 demonstrated high consistency with official agricultural statistics across all 11 Southeast Asian countries. As illustrated in Fig. 8, the Pearson correlation coefficients (*r*) consistently ranged from 0.90 to 0.95 for the years 2019–2023. This strong linear agreement confirms that our remote sensing-derived estimates effectively capture the magnitude and spatial distribution of agricultural activities at the national scale, aligning closely with the authoritative FAO cropland baseline. The persistent high correlation across multiple years further underscores the temporal stability of our mapping framework. Despite the fundamental methodological differences between bottom-up satellite observation and top-down census reporting, the convergence of these two independent datasets suggests that SEA_Cropland10 provides a reliable and spatially explicit alternative to traditional statistics. This reliability is particularly significant given the complexity of the Southeast Asian landscape, demonstrating that the dataset is robust enough to support regional-scale agricultural monitoring and policy assessment.

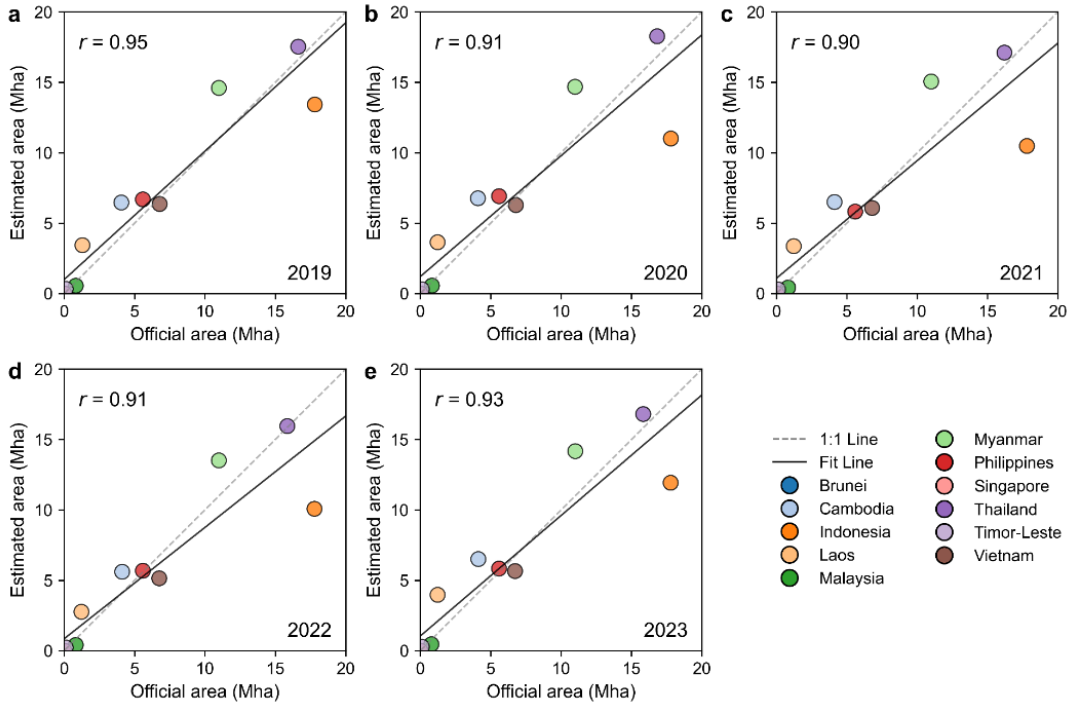


Figure 8: Comparison between the estimated cropland areas derived from this study and official agricultural statistics for Southeast Asian countries from 2019 to 2023 (a–e).

Stratified accuracy assessment using independent validation samples provides the mechanical explanation for these massive area discrepancies (Table 6). While global products maintain moderate sensitivity in flat regions (PA ≈ 77% for both GLAD and WorldCereal), their detection capabilities precipitously decline in sloping terrain. The PA for sloping cropland plummets to 32.86% (GLAD) and 29.67% (WorldCereal), indicating that approximately 70% of agricultural activities in marginal lands are systematically omitted by these global baselines. Conversely, their remarkably high UA (>94%) in these zones

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

reveal a highly conservative mapping strategy: existing algorithms prioritize minimizing false positives (commission errors) at the cost of severe omission errors. This trade-off renders them unsuitable for monitoring the fragmented expansion of highland agriculture, further validating the superior sensitivity of SEA_Cropland10 in complex terrains.

Table 6. Map accuracy of existing global product using stratified random samples.

Dataset	Metrics	Flat		Sloping		OA (%)
		Cropland	Non-cropland	Cropland	Non-cropland	
GLAD (2019)	PA (%)	77.37±4.33	92.13±3.31	32.86±6.31	98.93±1.04	79.33±2.29
	UA (%)	93.27±2.85	74.29±4.83	94.59±5.15	72.18±3.87	
WorldCereal (2021)	PA (%)	76.90±4.47	94.07±2.82	29.67±6.64	99.51±0.68	81.25±2.21
	UA (%)	94.27±2.73	76.28±4.57	96.43±4.86	75.94±3.63	

To assess the reliability of the SEA_Cropland10 dataset, we compared its area estimates with two global land-cover products (GLAD and WorldCereal) and the official agricultural statistics from FAOSTAT (Table 7). In terms of total cropland area, SEA_Cropland10 shows high consistency with the FAO statistics. For the year 2019, our dataset identified 69.4 Mha of cropland, slightly higher than but comparable to the FAO report of 64.1 Mha. In contrast, existing global products significantly underestimate the cropland extent in SEA. GLAD (2019) and WorldCereal (2021) detected only 54.2 Mha and 48.3 Mha, respectively, representing a substantial omission of approximately 15%–24% compared to the official statistics.

Table 7. Comparison of map-based sloping cropland area (>5°) across different datasets.

Dataset	Year	Flat cropland (Mha)	Sloping cropland (Mha)	Total Area (Mha)	Sloping ratio
SEA_Cropland10	2019	51.6	17.8	69.4	25.6%
SEA_Cropland10	2021	49.6	15.5	65.1	23.8%
GLAD	2019	48.1	6.1	54.2	11.3%
WorldCereal	2021	44.3	4.0	48.3	8.3%
FAOSTAT	2019	—	—	64.1	—
FAOSTAT	2021	—	—	63.6	—

A detailed breakdown by topographic slope reveals that the primary discrepancy stems from the detection of sloping cropland (>5°). Existing datasets perform adequately in flat regions but largely fail to capture agricultural activities in mountainous terrain. As shown in Table 7, for the year 2019, GLAD identified only 6.1 Mha of sloping cropland. Similarly, for 2021, WorldCereal detected merely 4.0 Mha. In sharp contrast, SEA_Cropland10 successfully mapped 17.8 Mha (2019) and 15.5 Mha (2021) of sloping cropland, capturing nearly 3 to 4 times the area identified by the global products. These results indicate that while global datasets like GLAD and WorldCereal are effective for large-scale, flat agricultural plains, they lack the sensitivity to detect the fragmented, small-scale farming systems prevalent in the Southeast Asian highlands.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

SEA_Cropland10 fills this critical data gap, providing a much more complete representation of the region's agricultural landscape, particularly in challenging sloping terrains.

A detailed visual comparison across six representative landscapes (Fig. 9) highlights the critical advantages of SEA_Cropland10 over GCEP30, GLAD, and WorldCereal. In general scenarios such as flat croplands (Column 1) and heterogeneous fragmented mosaics (Column 4), all products demonstrated reasonable detection capabilities; however, SEA_Cropland10 leveraged its 10-m resolution to resolve finer field boundaries and small patches that were often blurred in the 30-m GCEP30 and GLAD products. The performance gap widened significantly in challenging terrains. For sloping maize fields in Thailand (Column 2), terraces (Column 5), and mountainous agriculture in Laos (Column 6), global products exhibited varying degrees of omission. Notably, WorldCereal, despite sharing the high geometric fidelity of our product in plains, failed to detect the majority of sloping and terraced croplands. GCEP30 proved more robust than WorldCereal in these areas but still lacked completeness, while GLAD showed balanced but partial detection capabilities. Furthermore, regarding the unique uphill agricultural expansion on Indonesian volcanic slopes (Column 3), SEA_Cropland10 was the only dataset to accurately capture these high-altitude encroachment patterns, whereas other products largely missed them. Overall, SEA_Cropland10 effectively combines the high sensitivity required for marginal lands with the geometric precision of high-resolution mapping, outperforming existing products in complex tropical environments.

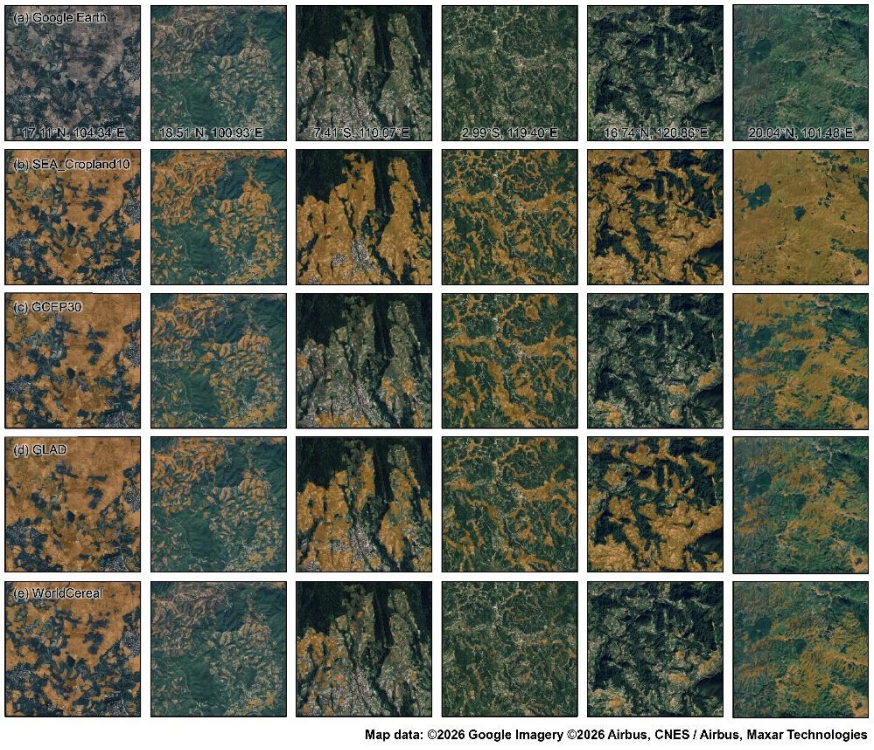


Figure 9: Visual comparison of the SEA_Cropland10 dataset with high-resolution satellite imagery and existing global cropland products across representative agricultural landscapes in Southeast Asia. Rows (**a**–**e**) correspond to different data sources: (**a**) Google Earth high-resolution satellite imagery (©2026 Google Imagery ©2026 Airbus, CNES / Airbus, Maxar Technologies); (**b**)

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

SEA_Cropland10 (this study, 10 m, 2024); (**c**) GCEP30 (30 m, 2015); (**d**) GLAD (30 m, 2019); and (**e**) WorldCereal (10 m, 2021). The columns illustrate distinct challenging scenarios. Orange pixels indicate identified cropland.

4.5 Limitations and perspectives

While our framework demonstrates robust performance in mapping complex agricultural landscapes, it entails an inherent methodological trade-off. To maximize the detection of fragmented, shifting, and sloping croplands, which are typologies frequently omitted in existing global products due to conservative mapping strategies, we prioritized high sensitivity (PA). This prioritization significantly reduces omission errors in marginal lands, albeit at the cost of an inevitable increase in commission errors within heterogeneous environments.

During the early establishment phase, tree crops exhibit spectral and textural signatures highly similar to those of cassava or maize (Maskell *et al.*, 2021; Wang *et al.*, 2023). Although the integration of AlphaEarth embeddings reduced these confusions compared to the baseline, some misclassification inevitably persists, particularly in Indonesia where plantation expansion is rapid. Future iterations could incorporate LiDAR-derived canopy height metrics (e.g., from GEDI or ICESat-2) or longer dense time-series analysis to better distinguish the structural evolution of tree crops from seasonal agricultural cycles (Adrah *et al.*, 2025).

Spatiotemporal discontinuity in optical observations limits phenological reconstruction. Although quarterly compositing and Sentinel-1 SAR integration were employed to mitigate tropical monsoon effects, persistent cloud cover resulted in data sparsity in specific regions. While interpolation techniques were applied to fill gaps, these reconstructed time series may lack the phenological fidelity required to fully capture high-frequency vegetation dynamics, thereby limiting the discriminative power between complex cropping systems and natural regrowth (Cai *et al.*, 2019).

The 10-m resolution of Sentinel imagery, while a significant upgrade from global baselines, still faces mixed-pixel limitations in ultra-fragmented landscapes. In parts of insular SEA, subsistence plots can be smaller than 0.05 ha or intercropped under sparse canopy coverage (Fritz *et al.*, 2015; Lesiv *et al.*, 2019; Waldner *et al.*, 2015). Our 0.1 ha MMU strategy, while necessary for reducing salt-and-pepper noise, may conservatively omit these micro-scale farming activities. Super-resolution techniques or the integration of sub-meter commercial imagery (e.g., PlanetScope) in a multi-stage sampling framework represents a viable pathway to capture this invisible cropland (Jia *et al.*, 2025; Rufin *et al.*, 2022).

From a methodological perspective, this study highlights the distinct value of GFM in data-sparse regions. Contrasting with recent findings in the U.S. where AlphaEarth showed limited gains over optimized baselines (Ma *et al.*, 2025), our results indicate that in the cloudy, noisy environments of SEA, these semantic embeddings are crucial for compensating for signal degradation. However, a critical bottleneck remains the physics-data gap inherent to current architectures. As purely data-driven systems, these models often lack physical constraints, making them susceptible to learning spurious correlations, such as confusing transient atmospheric noise with surface changes, rather than capturing the intrinsic biological consistency of crop phenology (Zhu *et al.*, 2026). Future research should focus on developing Time-Aware Foundation Models that can

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

natively ingest time-series data. Such advancement would unlock the ability to directly model complex phenological stages without relying on handcrafted temporal features (Hashemi *et al.*, 2025).

Beyond methodological advances, the SEA_Cropland10 dataset addresses a critical data gap in characterizing interactions between land use change and its environmental consequences across Southeast Asia. Robust quantification of fine-scale and temporally dynamic cropland expansion, particularly its progression onto topographically and ecologically sensitive slopes, remains limited, yet is fundamental for examining agriculture-related deforestation, soil erosion, hydrological alteration, and associated ecological processes in smallholder-dominated landscapes. By providing the first high-resolution of cropland dynamics spanning 2019–2024, SEA_Cropland10 establishes a consistent and reliable empirical baseline for future analyses. The dataset supports quantitative analyses of agriculture-driven land-use change and its associated climatic and ecohydrological consequences, and underpins future evaluations of smallholder system vulnerability and local sustainable livelihood interventions. In addition, it provides an essential data basis for characterizing and modeling the frequency and intensity of shifting agriculture practices across Southeast Asia. Collectively, SEA_Cropland10 offers a foundational data resource to support future efforts aimed at aligning agricultural development with long-term environmental sustainability in the region.

While SEA_Cropland10 establishes a robust modern baseline (2019–2024), understanding the long-term drivers of land-use change requires extending this record backward. A critical bottleneck is the incompatibility of modern foundation model embeddings with the Landsat archive (1980s–2010s). Developing cross-sensor transfer learning techniques, which can distill the knowledge learned from modern Sentinel/AlphaEarth models and adapt it to lower-resolution historical Landsat imagery, will be the key to reconstructing the multi-decadal history of agricultural expansion and shifting cultivation in the tropics.

5 Data availability

The SEA-Cropland10 is publicly available via Zenodo: https://doi.org/10.5281/zenodo.17828801 (Cai and Zeng, 2026) and Google Earth Engine App: https://ee-caiyt33tc.projects.earthengine.app/view/seacrop10 (last access: 9 January 2026).

6 Conclusions

We produced SEA_Cropland10, the first high-resolution (10-m) annual cropland dataset for SEA (2019–2024), by synergizing Sentinel-1/2 time series with AlphaEarth foundation model embeddings. This integration effectively compensated for signal degradation in the cloudy tropics, improving classification accuracy by 5.81%. The dataset achieves a robust multi-year overall accuracy of 92.67% (±1.47%) and high consistency with national statistics ($r > 0.90$). A breakthrough of this study is the recovery of sloping cropland (>5°) information. While global baselines historically overlook these marginal lands (PA = ~30%), SEA_Cropland10 elevates detection to 89–99%, identifying 3–4 times more sloping cropland area than previously reported. This reveals that agricultural expansion into erosion-prone terrain is far more

extensive than currently assessed. Based on unbiased stratified estimation, the total cropland area in SEA fluctuated from 68.7 (±2.8) Mha in 2019 to 67.3 (±3.2) Mha in 2024. By filling a critical data gap, SEA_Cropland10 offers a spatially detailed baseline for regional food security monitoring, carbon modeling, and sustainable land management in complex smallholder systems.

Author contributions

YC and ZZ designed the research. YC conducted the research and wrote the draft of the paper. ZZ supervised the research. All co-authors reviewed and revised the paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to thank the Google Earth Engine Team for supporting this research.

Financial support

This study is supported by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (grant no. JYB2025XDXM909), International Cooperation and Exchange of the National Natural Science Foundation of China (grant no. 42361144001), Shenzhen Science and Technology Project for Sustainable Development in Special Innovation (grant no. KCXFZ20230731093403008), SUSTech High Level of Special Funds (grants no. G03034K001, G030290001), the funding agencies of Zhejiang Province and Ningbo Municipality through the program "Novel Technologies for Joint Pollution Reduction and Carbon Sequestration", the State Key Laboratory of Climate Resilience for Coastal Cities and Otto Poon Research Institute for Climate-Resilient Infrastructure at the Hong Kong Polytechnic University. We acknowledge the Center for Computational Science and Engineering at the Southern University of Science and Technology for providing computing resources.

References

Adrah, E., Wong, J. P., and Yin, H.: Integrating GEDI, Sentinel-2, and Sentinel-1 imagery for tree crops mapping, Remote Sens. Environ., 319, 114644, https://doi.org/10.1016/j.rse.2025.114644, 2025.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Bhogapurapu, N., Dey, S., Mandal, D., Bhattacharya, A., Karthikeyan, L., McNairn, H., and Rao, Y. S.: Soil moisture retrieval over croplands using dual-pol L-band GRD SAR data, Remote Sens. Environ., 271, 112900, https://doi.org/10.1016/j.rse.2022.112900, 2022.

Breiman, L.: Classification and Regression Trees, Routledge, New York, 368 pp., https://doi.org/10.1201/9781315139470, 2017.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., and Tait, A. M.: Dynamic World, Near real-time global 10 m land use land cover mapping, Sci. Data, 9, 251, https://doi.org/10.1038/s41597-022-01307-4, 2022.

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., and Kohli, P.: AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data, arXiv [preprint], https://doi.org/10.48550/arXiv.2507.22291, 8 September 2025.

Cai, Y. and Zeng, Z.: SEA_Cropland10: An Annual Cropland Dataset for Southeast Asia at 10-m Resolution since 2019, Zenodo [data set], https://doi.org/10.5281/zenodo.17828801, 2026.

Cai, Y., Lin, H., and Zhang, M.: Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data, Adv. Space Res., 64, 2233–2244, https://doi.org/10.1016/j.asr.2019.08.042, 2019.

Cai, Y., Xu, X., Zhu, P., Nie, S., Wang, C., Xiong, Y., and Liu, X.: Unveiling spatiotemporal tree cover patterns in China: The first 30 m annual tree cover mapping from 1985 to 2023, ISPRS J. Photogramm. Remote Sens., 216, 240–258, https://doi.org/10.1016/j.isprsjprs.2024.08.001, 2024.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J.: Global land cover mapping at 30m resolution: A POK-based operational approach, ISPRS J. Photogramm. Remote Sens., 103, 7–27, https://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Cochran, W. G.: Sampling techniques, Wiley, New York, NY, ISBN 9788126515240, 1977.

Eilers, P. H. C.: A Perfect Smoother, Anal. Chem., 75, 3631–3636, https://doi.org/10.1021/ac034173t, 2003.

ESA: Land Cover CCI Product User Guide Version 2, European Space Agency, Tech. Rep. 2017.

FAO (Ed.): Guidelines for soil description, 4. ed., Food and Agriculture Organization of the United Nations, Rome, 95 pp., 2006.

FAO: FAOSTAT Statistical Database, Food and Agriculture Organization of the United Nations, Rome, Italy, available at: https://www.fao.org/faostat/en/ (last access: 19 January 2026), 2025.

FAOSTAT: Crops and livestock products, Food and Agriculture Organization of the United Nations, Rome, Italy, 2023.

Feng, Y., Ziegler, A. D., Elsen, P. R., Liu, Y., He, X., Spracklen, D. V., Holden, J., Jiang, X., Zheng, C., and Zeng, Z.: Upward expansion and acceleration of forest clearance in the mountains of Southeast Asia, Nat. Sustain., 4, 892–899, https://doi.org/10.1038/s41893-021-00738-y, 2021.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Feng, Y., Zeng, Z., Searchinger, T. D., Ziegler, A. D., Wu, J., Wang, D., He, X., Elsen, P. R., Ciais, P., Xu, R., Guo, Z., Peng, L., Tao, Y., Spracklen, D. V., Holden, J., Liu, X., Zheng, Y., Xu, P., Chen, J., Jiang, X., Song, X.-P., Lakshmi, V., Wood, E. F., and Zheng, C.: Doubling of annual forest carbon loss over the tropics during the early twenty-first century, Nat. Sustain., 5, 444–451, https://doi.org/10.1038/s41893-022-00854-3, 2022.

Feng, Z., Atzberger, C., Jaffer, S., Knezevic, J., Sormunen, S., Young, R., Lisaius, M. C., Immitzer, M., Jackson, T., Ball, J., Coomes, D. A., Madhavapeddy, A., Blake, A., and Keshav, S.: TESSERA: Temporal Embeddings of Surface Spectra for Earth Representation and Analysis, arXiv [preprint], https://doi.org/10.48550/arXiv.2506.20380, 19 November 2025.

Friedl, M. and Sulla-Menashe, D.: MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061, https://doi.org/10.5067/MODIS/MCD12Q1.061, 2022.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Abdel Aziz, S., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D. R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., and Obersteiner, M.: Mapping global cropland and field size, Glob. Change Biol., 21, 1980–1992, https://doi.org/10.1111/gcb.12838, 2015.

Gregorio, A. D.: Land Cover Classification System: Classification Concepts and User Manual: LCCS, Food and Agriculture Organization, 212 pp., ISBN: 9251053278, 2005.

Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Cheng, F., Zhuang, H., Zhang, J., and Tao, F.: NESEA-Rice10: high-resolution annual paddy rice maps for Northeast and Southeast Asia from 2017 to 2019, Earth Syst. Sci. Data, 13, 5969–5986, https://doi.org/10.5194/essd-13-5969-2021, 2021.

Jia, X., Hao, Z., Sun, L., Yang, Q., Wang, Z., Yin, Z., Shi, L., Li, X., Du, Y., and Ling, F.: Super-Resolution Cropland Mapping by Spectral and Spatial Training Samples Simulation, IEEE Trans. Geosci. Remote Sens., 63, 1–15, https://doi.org/10.1109/TGRS.2025.3547042, 2025.

Jung, M., Arnell, A., de Lamo, X., García-Rangel, S., Lewis, M., Mark, J., Merow, C., Miles, L., Ondo, I., Pironon, S., Ravilious, C., Rivers, M., Schepaschenko, D., Tallowin, O., van Soesbergen, A., Govaerts, R., Boyle, B. L., Enquist, B. J., Feng, X., Gallagher, R., Maitner, B., Meiri, S., Mulligan, M., Ofer, G., Roll, U., Hanson, J. O., Jetz, W., Di Marco, M., McGowan, J., Rinnan, D. S., Sachs, J. D., Lesiv, M., Adams, V. M., Andrew, S. C., Burger, J. R., Hannah, L., Marquet, P. A., McCarthy, J. K., Morueta-Holme, N., Newman, E. A., Park, D. S., Roehrdanz, P. R., Svenning, J.-C., Violle, C., Wieringa, J. J., Wynne, G., Fritz, S., Strassburg, B. B. N., Obersteiner, M., Kapos, V., Burgess, N., Schmidt-Traub, G., and Visconti, P.: Areas of global importance for conserving terrestrial biodiversity, carbon and water, Nat. Ecol. Evol., 5, 1499–1509, https://doi.org/10.1038/s41559-021-01528-7, 2021.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., and Brumby, S. P.: Global land use / land cover with Sentinel 2 and deep learning, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021

IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 12–16 July 2021, 4704–4707, https://doi.org/10.1109/IGARSS47720.2021.9553499, 2021.

Lesiv, M., Laso Bayas, J. C., See, L., Duerauer, M., Dahlia, D., Durando, N., Hazarika, R., Kumar Sahariah, P., Vakolyuk, M., Blyshchyk, V., Bilous, A., Perez-Hoyos, A., Gengler, S., Prestele, R., Bilous, S., Akhtar, I. ul H., Singha, K., Choudhury, S. B., Chetri, T., Malek, Ž., Bungnamei, K., Saikia, A., Sahariah, D., Narzary, W., Danylo, O., Sturn, T., Karner, M., McCallum, I., Schepaschenko, D., Moltchanova, E., Fraisl, D., Moorthy, I., and Fritz, S.: Estimating the global distribution of field size using crowdsourcing, Glob. Change Biol., 25, 174–186, https://doi.org/10.1111/gcb.14492, 2019.

Li, P., Feng, Z., Jiang, L., Liao, C., and Zhang, J.: A Review of Swidden Agriculture in Southeast Asia, Remote Sens., 6, 1654–1683, https://doi.org/10.3390/rs6021654, 2014.

Li, W., Liang, S., Chen, K., Chen, Y., Ma, H., Xu, J., Ma, Y., Zhang, Y., Guan, S., Fang, H., and Shi, Z.: AgriFM: A multi-source temporal remote sensing foundation model for Agriculture mapping, Remote Sens. Environ., 334, 115234, https://doi.org/10.1016/j.rse.2026.115234, 2026.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, NIPS, California, United States, 4–9 December 2017, 2017.

Ma, Y., Shen, Y., Swatantran, A., and Lobell, D. B.: Harvesting AlphaEarth: Benchmarking the Geospatial Foundation Model for Agricultural Downstream Tasks, arXiv [preprint], https://doi.org/10.48550/arXiv.2601.00857, 30 December 2025.

Maskell, G., Chemura, A., Nguyen, H., Gornott, C., and Mondal, P.: Integration of Sentinel optical and radar data for mapping smallholder coffee production systems in Vietnam, Remote Sens. Environ., 266, 112709, https://doi.org/10.1016/j.rse.2021.112709, 2021.

Miettinen, J., Stibig, H.-J., and Achard, F.: Remote sensing of forest degradation in Southeast Asia—Aiming for a regional view through 5–30 m satellite data, Glob. Ecol. Conserv., 2, 24–36, https://doi.org/10.1016/j.gecco.2014.07.007, 2014.

Miettinen, J., Hooijer, A., Vernimmen, R., Liew, S. C., and Page, S. E.: From carbon sink to carbon source: extensive peat oxidation in insular Southeast Asia since 1990, Environ. Res. Lett., 12, 024014, https://doi.org/10.1088/1748-9326/aa5b6f, 2017.

Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N., and Reiche, J.: Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine, Remote Sens., 13, 1954, https://doi.org/10.3390/rs13101954, 2021.

Oliphant, A. J., Thenkabail, P. S., Teluguntla, P., Xiong, J., Gumma, M. K., Congalton, R. G., and Yadav, K.: Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud, Int. J. Appl. Earth Obs. Geoinformation, 81, 110–124, https://doi.org/10.1016/j.jag.2018.11.014, 2019.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sens. Environ., 148, 42–57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Parker, D., Tosiani, A., Yazid, M., Sari, I. L., Kartika, T., Kustiyo, N., Firmansyah, R., Said, Z., Wijaya, A., Potapov, P., Tyukavina, A., Stehman, S. V., Zalles, V., Pickens, A., Pickering, J., Turubanova, S., and Hansen, M. C.: Land in limbo: Nearly one third of Indonesia's cleared old-growth forests left idle, Proc. Natl. Acad. Sci., 121, e2318029121, https://doi.org/10.1073/pnas.2318029121, 2024.

Pasquarella, V. J., Brown, C. F., Czerwinski, W., and Rucklidge, W. J.: Comprehensive quality assessment of optical satellite imagery using weakly supervised video learning, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2125–2135, https://doi.org/10.1109/CVPRW59228.2023.00206, 2023.

Pérez-Hoyos, A., Rembold, F., Kerdiles, H., and Gallego, J.: Comparison of Global Land Cover Datasets for Cropland Monitoring, Remote Sens., 9, 1118, https://doi.org/10.3390/rs9111118, 2017.

Pipia, L., Muñoz-Marí, J., Amin, E., Belda, S., Camps-Valls, G., and Verrelst, J.: Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes, Remote Sens. Environ., 235, 111452, https://doi.org/10.1016/j.rse.2019.111452, 2019.

Potapov, P., Turubanova, S., Hansen, M. C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., and Cortez, J.: Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century, Nat. Food, 3, 19–28, https://doi.org/10.1038/s43016-021-00429-z, 2022.

Pravalprukskul, P., Bruun, T. B., and Messerli, P.: Maize boom, bust and beyond: Investigating land use transitions in the northern Thai uplands, Land Use Policy, 132, 106815, https://doi.org/10.1016/j.landusepol.2023.106815, 2023.

Rufin, P., Bey, A., Picoli, M., and Meyfroidt, P.: Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data, Int. J. Appl. Earth Obs. Geoinformation, 112, 102937, https://doi.org/10.1016/j.jag.2022.102937, 2022.

Struebig, M. J., Lee, J. S. H., Deere, N. J., Gevaña, D. T., Ingram, D. J., Lwin, N., Nguyen, T., Santika, T., Seaman, D. J. I., Supriatna, J., and Davies, Z. G.: Drivers and solutions to Southeast Asia's biodiversity crisis, Nat. Rev. Biodivers., 1, 497–514, https://doi.org/10.1038/s44358-025-00064-7, 2025.

Sun, C., Zhang, H., Xu, L., Ge, J., Jiang, J., Zuo, L., and Wang, C.: Twenty-meter annual paddy rice area map for mainland Southeast Asia using Sentinel-1 synthetic-aperture-radar data, Earth Syst. Sci. Data, 15, 1501–1520, https://doi.org/10.5194/essd-15-1501-2023, 2023.

Thenkabail, P. S., Teluguntla, P. G., Xiong, J., Oliphant, A., Congalton, R. G., Ozdogan, M., Gumma, M. K., Tilton, J. C., Giri, C., Milesi, C., Phalke, A., Massey, R., Yadav, K., Sankey, T., Zhong, Y., Aneece, I., and Foley, D.: Global cropland-extent product at 30-m resolution (GCEP30) derived from Landsat satellite time-series data for the year 2015 using multiple machine-learning algorithms on Google Earth Engine cloud, Prof. Pap., 63, https://doi.org/10.3133/pp1868, 2021.

Tubiello, F. N., Conchedda, G., Casse, L., Hao, P., De Santis, G., and Chen, Z.: A new cropland area database by country circa 2020, Earth Syst. Sci. Data, 15, 4997–5015, https://doi.org/10.5194/essd-15-4997-2023, 2023.

https://doi.org/10.5194/essd-2026-78
Preprint. Discussion started: 25 February 2026

Valentin, C., Agus, F., Alamban, R., Boosaner, A., Bricquet, J. P., Chaplot, V., de Guzman, T., de Rouw, A., Janeau, J. L., Orange, D., Phachomphonh, K., Do Duy Phai, Podwojewski, P., Ribolzi, O., Silvera, N., Subagyono, K., Thiébaux, J. P., Tran Duc Toan, and Vadari, T.: Runoff and sediment losses from 27 upland catchments in Southeast Asia: Impact of rapid land use changes and conservation practices, Agric. Ecosyst. Environ., 128, 225–238, https://doi.org/10.1016/j.agee.2008.06.004, 2008.

Van Tricht, K., Degerickx, J., Gilliams, S., Zanaga, D., Battude, M., Grosu, A., Brombacher, J., Lesiv, M., Bayas, J. C. L., Karanam, S., Fritz, S., Becker-Reshef, I., Franch, B., Mollà-Bononad, B., Boogaard, H., Pratihast, A. K., Koetz, B., and Szantoi, Z.: WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping, Earth Syst. Sci. Data, 15, 5491–5515, https://doi.org/10.5194/essd-15-5491-2023, 2023.

Waldner, F., Fritz, S., Di Gregorio, A., and Defourny, P.: Mapping Priorities to Focus Cropland Mapping Activities: Fitness Assessment of Existing Global, Regional and National Cropland Maps, Remote Sens., 7, 7959–7986, https://doi.org/10.3390/rs70607959, 2015.

Wang, Y., Hollingsworth, P. M., Zhai, D., West, C. D., Green, J. M. H., Chen, H., Hurni, K., Su, Y., Warren-Thomas, E., Xu, J., and Ahrends, A.: High-resolution maps show that rubber causes substantial deforestation, Nature, 623, 340–346, https://doi.org/10.1038/s41586-023-06642-z, 2023.

World Bank: World Development Indicators, The World Bank Group, Washington, D.C., USA, available at: https://databank.worldbank.org/source/world-development-indicators (last access: 19 January 2026), 2025.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., and Arino, O.: ESA WorldCover 10 m 2021 v200, Zenodo [data set] https://doi.org/10.5281/zenodo.7254221, 2022.

Zeng, Z., Gower, D. B., and Wood, E. F.: Accelerating forest loss in Southeast Asian Massif in the 21st century: A case study in Nan Province, Thailand, Glob. Change Biol., 24, 4682–4695, https://doi.org/10.1111/gcb.14366, 2018a.

Zeng, Z., Estes, L., Ziegler, A. D., Chen, A., Searchinger, T., Hua, F., Guan, K., Jintrawet, A., and F. Wood, E.: Highland cropland expansion and forest loss in Southeast Asia in the twenty-first century, Nat. Geosci., 11, 556–562, https://doi.org/10.1038/s41561-018-0166-9, 2018b.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J.: GLC_FCS30: global land-cover product with fine classification system at 30 m using time-series Landsat imagery, Earth Syst. Sci. Data, 13, 2753–2776, https://doi.org/10.5194/essd-13-2753-2021, 2021.

Zhu, X. X., Xiong, Z., Wang, Y., Stewart, A. J., Heidler, K., Wang, Y., Yuan, Z., Dujardin, T., Xu, Q., and Shi, Y.: On the foundations of Earth foundation models, Commun. Earth Environ., https://doi.org/10.1038/s43247-025-03127-x, 2026.