https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Open Access Earth System Science Data Discussions

Global 30-m annual cropland extent dynamics (2000–2024): A consistent baseline of structural evolution and regional disparities

Yuanhong Liao[1,2,†], Shuang Chen[3,†], Yuqi Bai[1,2,*], Jie Wang[4,*], Peng Gong[3,5,*]

[1]Department of Earth System Science, Institute for Global Change Studies, Ministry of Education Ecological Field Station for East Asian Migratory Birds, Tsinghua University, Beijing 100084, China
[2]Tsinghua Urban Institute, Tsinghua University, Beijing 100084, China
[3]Department of Geography, The University of Hong Kong, Hong Kong, China
[4]Pengcheng Laboratory, Shenzhen 518000, China
[5] Institute for Climate and Carbon Neutrality and Department of Earth Sciences, The University of Hong Kong, Hong Kong, China

Correspondence to: Yuqi Bai (yuqibai@gmail.com), Jie Wang (wangj10@pcl.ac.cn), Peng Gong (penggong@hku.hk)
[†]: Y. Liao and S. Chen contributed equally to this work

Abstract. Accurate quantification of the structural evolution of global agricultural systems is critical for assessing food security and monitoring planetary boundaries. However, current agricultural monitoring is hindered by a baseline that reflects the continuous, annual nature of agricultural management and distinguishes active cultivation cycles from permanent structural changes. Existing products typically rely on fragmented snapshots, multi-year aggregation, or generalized and inconsistent definitions (introducing systematic bias and noise into trend analysis). To bridge this gap, we generated the first Global 30-m Annual Cropland Extent Dynamics (GACED30) dataset (2000–2024). Our continuous mapping framework integrates the gap-free SDC30 to capture the intra-annual phenological transitions distinctive of active cultivation, employs a spectral-semantic sample alignment strategy to resolve the inconsistencies across different cropland sample sets, and applies a rule-based processing to mitigate the spectral ambiguity between active fallow and natural bareland. Furthermore, we derived grid-based Structural Evolution Indicators via continuous statistical trend analysis, utilizing the full 25-year time-series density to rigorously quantify expansion and abandonment trajectories. Comprehensive assessment demonstrates the reliability of GACED30, which achieves a high overall accuracy of 96.5% and significantly outperforms existing global products (e.g., GLAD Cropland, ESA World Cover and GLC_FCS30D) in capturing temporal stability. Crucially, GACED30 exhibits strong agreement with FAO national statistics, achieving a high correlation in cropland area (R^2: 0.95) and an 81.1% consistency in production trends. Based on this consistent baseline, we analyzed the structural evolution of global agriculture, estimating the 2024 cropland area at 1488.5 Mha with a net change rate of 1.2 Mha/year. Our trend analysis reveals a distinct global divergence: structural expansion is heavily concentrated in the Global South (e.g., Africa and South America), driven by commodity frontiers, whereas the Global North is characterized by widespread stability or policy-driven contraction. GACED30 thus provides a reliable evidence base for monitoring the changing footprint of global agriculture. The dataset is publicly available at https://doi.org/10.5281/zenodo.18199675 (Chen et al., 2026).

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

1 Introduction

Ensuring global food security while remaining within planetary boundaries is one of the defining challenges of the 21st century (Godfray et al., 2010; FAO, 2017). As demographic pressures and climate volatility escalate the demand for food (Tilman et al., 2011), the global agricultural sector faces a critical "land dilemma": the imperative to increase production versus the urgent need to halt the encroachment of cropland into carbon-rich ecosystems and biodiversity hotspots (Foley et al., 2011; Zhang et al., 2025b; Ceaușu et al., 2025). Resolving this trade-off requires a paradigm shift from static assessments of total acreage to a dynamic understanding of structural evolution, i.e., the long-term spatiotemporal trajectory of agricultural expansion, stability, and retraction (Potapov et al., 2022). Robust information on these long-term cropland dynamics is a foundational prerequisite for diagnosing the health of the global land system (Tu et al., 2024; FAO, 2025). By rigorously distinguishing between permanent land cover conversion (e.g., commodity-driven deforestation or urbanization) and the rotational variability inherent to management cycles, such data empowers policymakers to differentiate true sustainable intensification from extractive land reclamation (Tubiello et al., 2021; Pretty, 2018).

The progression of global cropland monitoring has been fundamentally transformed by the convergence of open satellite archives (Wulder et al., 2012; Drusch et al., 2012), advanced machine learning (Zhu et al., 2017), and cloud computing (Gorelick et al., 2017), enabling the emergence of a new generation of high-resolution (10–30 m) mapping products (Table 1). However, despite the emergence of high-resolution mapping products (Table 1), current operational datasets exhibit critical limitations when applied to the monitoring of long-term structural evolution. These products generally fall into four categories, each constrained by specific trade-offs. The first two categories comprise short-term thematic studies (e.g., the 30-m GCEP30 (Thenkabail et al., 2021) and 10-m WorldCereal (Van Tricht et al., 2023)) and short-term comprehensive land cover products (e.g., ESA WorldCover (Zanaga et al., 2022), Esri Land Cover (Karra et al., 2021), and GLC_FCS10 (Zhang et al., 2025c)). While offering high spatial fidelity, these datasets function as independent temporal fragments or isolated snapshots. Lacking multi-decadal continuity, they cannot capture the historical trajectory required to distinguish current structural patterns from past land use dynamics. The third category involves long-term thematic products, represented by Global Land Analysis and Discovery (GLAD) Cropland (Potapov et al., 2022). While achieving high thematic accuracy, this dataset relies on multi-year epoch aggregation (e.g., 4-year composites) to mitigate observational gaps. Consequently, it primarily supports bi-temporal change detection (comparing aggregated epochs) rather than continuous annual monitoring. This aggregation strategy effectively blurs inter-annual dynamics, masking the exact timing of expansion or abandonment and limiting the ability to track detailed structural evolution. The fourth category consists of long-term comprehensive land cover products, such as GLC_FCS30D (Zhang et al., 2023). Although offering a multi-decadal span (1985–2022), these generalist products frequently adopt broad definitions that encompass rainfed, herbaceous, and shrub-based mosaics, leading to systematic area overestimation (Tubiello et al., 2023b). Furthermore, their classification systems often fail to distinguish fallow-rotation cycles from natural vegetation changes. This semantic ambiguity results in significant statistical fluctuations in cropland area, identifying noise rather than signal, and making it impossible to robustly judge real structural trends. Consequently, a critical gap remains for a global product that

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

simultaneously achieves the targeted accuracy of thematic maps and the annual, definitionally consistent baseline required to quantify the structural evolution of global agriculture.

Table 1: Summary of major global high-resolution (10 m and 30 m) land cover products relevant to cropland monitoring.

Product name	Product type	Cropland definition	Spatial scale	Temporal attributes	Input satellites	Validation protocol & Accuracy
WorldCereal (Van Tricht et al., 2023)	Thematic	Temporary crops only. Excludes perennial crops (orchards) and pastures.	10 m	Annual & seasonal products for 2021	Sentinel-1/2, Landsat 8	Global 50,000 samples (UA:0.89; PA:0.92)
WorldCover (Zanaga et al., 2022)	General	Annual, harvestable crops. Excludes perennial woody crops (e.g., orchards) and greenhouses.	10 m	Annual 2000-2001	Sentinel-1/2	Global 2,162,366 sample units (UA:0.81; PA:0.79)
GLC_FCS10 (Zhang et al., 2025c)	General	Include rainfed cropland, irrigated cropland, herbaceous cover cropland, and tree or shrub cover cropland.	10 m	Annual for 2023	Sentinel-1/2	Global 56,121 samples (UA:0.82; PA:0.88)
Esri Land Cover (Karra et al., 2021)	General	Human planted/plotted cereals, grasses, and crops not at tree height.	10 m	Annual 2017-2024	Sentinel-2	Global 24,000 sample patches
GLAD Cropland (Potapov et al., 2022)	Thematic	Annual and perennial herbaceous crops. Excludes tree crops (orchards) and permanent pastures.	30 m	4-years epochs for 2000-2019	Landsat (full archive)	Global 3,5000 samples for 2003 to 2019 (UA:0.89; PA:0.86)
GCEP30 (Thenkabail et al., 2021)	Thematic	Includes annual crops, fallow land, and permanent crops (orchards, plantations).	30 m	Annual for 2015	Landsat 7-8	Global 19,171 samples (UA:0.78; PA:0.83)
FROM-GC (Yu et al., 2013)	Thematic	Consistent with FAO's definition of 'arable land' plus 'permanent cropland'.	30 m	Annual 2010	Landsat & MODIS	Compare the cropland area with FAOSTAT at country level (R^2:0.97)
GLC_FCS30D (Zhang et al., 2023)	General	Include rainfed cropland, irrigated cropland, herbaceous cover cropland, and tree or shrub cover cropland.	30 m	Annual for 2000-2022 and 5-years epochs for 1985-2000	Landsat (full archive)	Global 84,526 samples for 2020 (UA:0.86; PA:0.87)

The primary observational barrier to annual cropland mapping is the requirement for continuous, gap-free time series to capture the distinct intra-annual phenological pulse of agriculture. Unlike stable land cover types, the reliable separation of active cultivation from spectrally similar natural vegetation (e.g., grasslands) depends entirely on detecting the temporal signature of planting, green-up, and senescence (Bégué et al., 2018; Zhang et al., 2025a). However, preserving this high-frequency signal in optical satellite imagery is severely hampered by cloud contamination, which creates critical data gaps (Whitcraft et al., 2015). While strategies such as multi-year epoch aggregation (Potapov et al., 2022) effectively mitigate

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

data gaps, they do so by sacrificing the annual temporal fidelity required to distinguish cropping cycles from natural variability. Consequently, recent advancements have shifted toward dense time-series reconstruction through sensor fusion—such as the Harmonized Landsat and Sentinel-2 (HLS) project (Claverie et al., 2018) and the generation of Seamless Data Cubes (SDC) (Chen et al., 2024; Liang et al., 2023; Liu et al., 2021).

While these technologies provide the prerequisite observational continuity for monitoring agricultural dynamics, the consistent observation of agricultural systems is equally constrained by inconsistencies in the labelling time and semantics of training samples. While the theory of stable classification with finite samples suggests that classifiers can tolerate a certain margin of random labelling error during sample transferring (Gong et al., 2019; Gong et al., 2024), it is ill-equipped to handle the systematic errors arising from transitional states—such as temporary grasslands or fallow periods—where semantic definitions of 'cropland' diverge. Besides these semantic inconsistencies, robust monitoring is further impeded by the inherent spectral ambiguity between active fallow phases and natural bareland, where the temporary absence of vegetation often leads to false negatives that induce spurious temporal instability and compromise the reliability of long-term trend estimations. Consequently, robust annual mapping necessitates a spectral-semantic alignment strategy and post-processing rules to rigorously filter these temporal and definition-level conflicts, unifying the spectral representations of diverse samples into a single, thematically consistent classification system.

To address the challenges, we generated the first Global 30-m Annual Cropland Extent Dynamics (GACED30) dataset spanning 2000 to 2024 by developing a continuous mapping framework that integrates the gap-free 30-m Seamless Data Cube (SDC30) with a spectral-semantic sample alignment strategy and a spatiotemporal refinement module. Comprehensive assessments indicate that GACED30 outperforms existing global products and achieves high consistency with Food and Agriculture Organization (FAO) national statistics. Leveraging this consistent baseline, we quantify the long-term structural evolution of global agriculture. We derived a Cropping Frequency layer (to characterize system stability) and a 1-km Structural Trend Indicator, which rigorously separates significant land use conversion (expansion/abandonment) from inter-annual phenological variability. This analysis reveals the distinct divergent agricultural trajectories between the expanding Global South and the stabilizing Global North, providing a reliable evidence base for monitoring the changing footprint of global food systems. The dataset is publicly available at https://doi.org/10.5281/zenodo.18199675 (Chen et al., 2026).

The remainder of this paper is organized as follows: Sect. 2 and Sect. 3 detail the datasets and the phenology-adaptive mapping framework; Sect. 4 presents the comprehensive validation results, harmonization with FAO statistics, and derived global cropping frequency and structural agricultural trends; and Sect. 5 discusses the product's advantages and limitations before concluding in Sect. 6.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Open Access Earth System Science Data Discussions

2 Datasets

The generation and validation of the GACED30 dataset relied on a multi-faceted data strategy. This involved leveraging a high-density satellite data cube for primary analysis, incorporating auxiliary datasets for contextual information, utilizing extensive sample sets for model training and validation, and referencing official agricultural statistics for area comparison.

2.1 Definition and harmonization with FAO standards

Unlike general land cover products that characterize the instantaneous physical state of the surface (e.g., bare soil or green vegetation) (Di Gregorio, 2005), GACED30 defines "Cropland" based on the evidence of active anthropogenic management cycles. Crucially, we view the temporal absence of vegetation not merely as "bare land," but often as a functional phase of the agricultural system—specifically, the fallow period required for soil recovery or rotation. To ensure the dataset serves as a valid baseline for food security assessment, we aligned our definition with the FAO standards, specifically targeting the Cropland category, which aggregates "Arable land" and "Permanent crops" (Tubiello et al., 2023b). Consequently, the GACED30 cropland class encompasses the following actively managed categories:

- Annual Herbaceous Crops: Corresponding to the FAO's "Temporary crops," this includes land used for crops with a less-than-one-year growing cycle, such as cereals (e.g., rice, wheat, maize) and vegetables.
- Perennial Woody Crops: Corresponding to FAO's "Permanent crops," this includes orchards, vineyards, and plantations that do not require replanting for several years. This inclusion is critical, as many existing remote sensing products exclude woody crops, leading to significant area underestimations.
- Active Fallow: Aligning with FAO's "Temporary fallow," we include land that is temporarily resting (bare farmland) as part of the cultivation rotation. By utilizing long-term phenological data, we distinguish this active fallow from abandoned land or natural barren surfaces.
- Agricultural Structures: This includes greenhouses and other permanent production structures essential to modern agricultural systems.

While FAO includes "Temporary meadows and pastures" (land cultivated with forage for less than five years), these areas are often spectrally indistinguishable from natural grasslands in satellite imagery. To avoid the systematic overestimation that conflate rangelands with cropland, GACED30 explicitly excludes temporary meadows and pastures.

2.2 Input datasets

The primary observational barrier to annual cropland mapping is the requirement for continuous time series to capture the distinct intra-annual phenological pulse of agriculture. Preserving this high-frequency signal in optical satellite imagery is severely hampered by cloud contamination, which creates critical data gaps (Sassen et al., 2008). This challenge is particularly acute in intensive agricultural zones within tropical and monsoonal regions—such as the Chocó Department in Colombia or the Ganges Delta—where cloud cover frequently obscures key phenological stages like sowing or peak

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

greenness. To resolve this, we utilized the Global 30-m Seamless Data Cube (Chen et al., 2023; Chen et al., 2024). Unlike standard composites that aggregate observations over long periods to minimize gaps, SDC30 provides a daily, gap-free surface reflectance record from 2000 to 2024. Its seamless consistency is critical for the distinctive temporal signature of cultivation—rapid green-up followed by harvest—can be reliably distinguished from natural vegetation phenology (Bégué et al., 2018; Zhang et al., 2025a), even in data-sparse regions where traditional optical baselines fail.

To further refine the semantic distinction between agricultural and natural systems, we integrated agro-ecological constraints derived from the NASADEM (NASA JPL, 2020). Beyond providing simple topographic context, these elevation and slope metrics serve a critical specific function in our framework: differentiating "active fallow" from "natural bare land." By identifying terrain with slope constraints unsuitable for machinery or cultivation, we can effectively filter out natural barren areas that mimic the spectral signature of fallow fields, ensuring that "bare" pixels are only classified as cropland within physiographically arable zones.

Further, the GLAD Cropland product served as a thematic stratification layer. By leveraging its long-term record (Potapov et al., 2022) as a spatial prior, we optimized the representativeness of our training samples, ensuring the model was primed with a diverse range of agricultural spectral signatures from established cultivation zones.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

2.3 Sample sets

2.3.1 Training samples

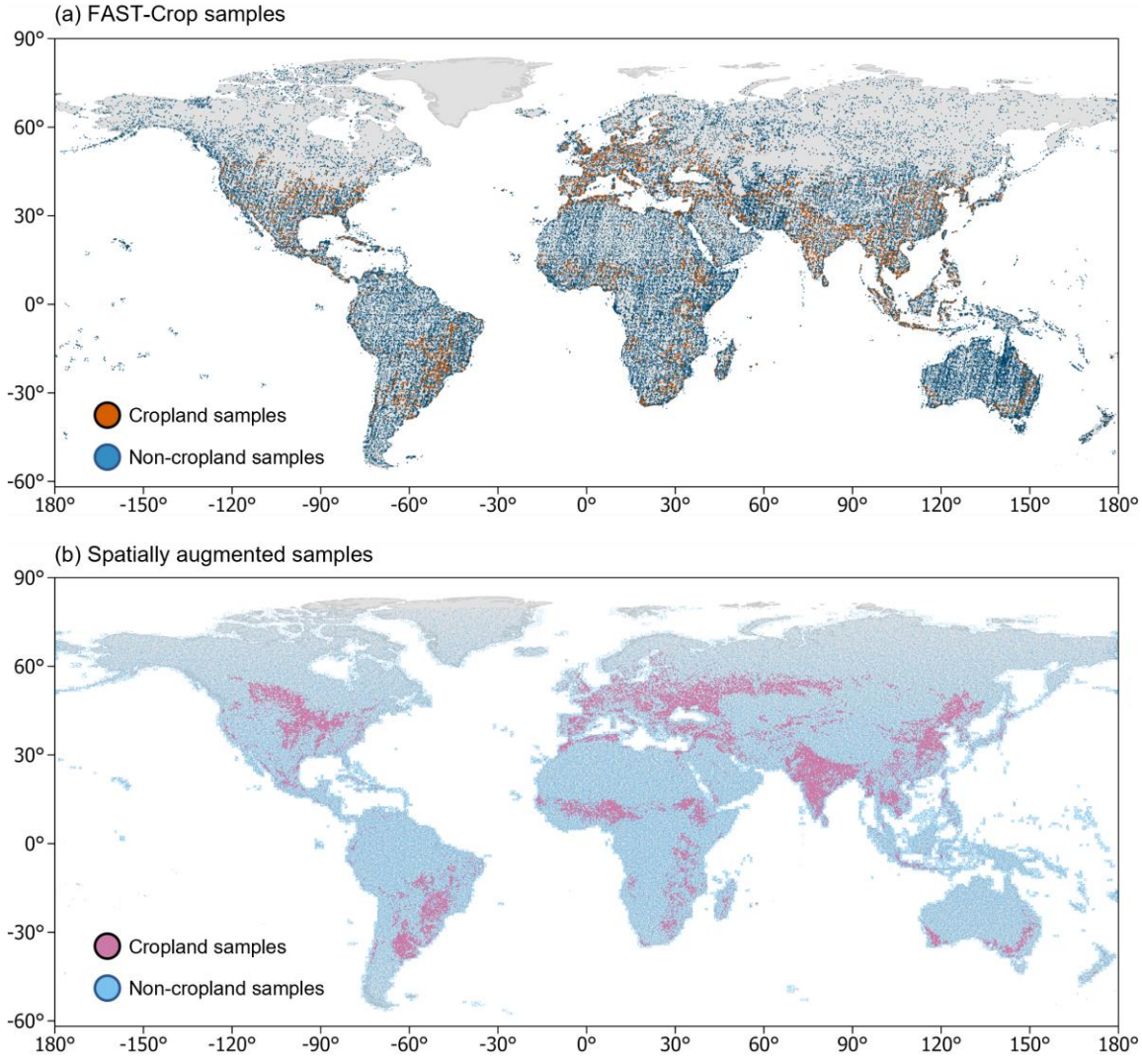


Figure 1. Spatial distribution of the training sample sets used for GACED30 generation. (a) FAST-Crop sample sets derived by expert annotation; (b) spatially augmented sample sets generated via rule-based expansion to improve global coverage.

To ensure high thematic fidelity across diverse agro-ecological zones, we constructed a composited training set (Fig. 1), which is composed of two distinct tiers of reference data: the high-confidence FAST-Crop sample set and the spatially augmented set. The foundation of our training strategy is based on the First All-Season Sample Set (FAST) (Li et al., 2017). While FAST provides a general land cover baseline, its value for this study lies in its rigorous, multi-seasonal annotation protocol. Expert interpreters utilized seasonal Landsat 8 imagery, high-resolution Google Earth images, and MODIS

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

phenology profiles to verify land cover stability. This "all-season" verification is essential for our specific definition of cropland (Sect. 2.1), as it allows us to confirm the presence of active management cycles (e.g., plowing, sowing, harvest) rather than relying on a single static snapshot. We filtered the original sample set to strictly select 66,194 samples (9,924 cropland and 56,270 non-cropland) that align with our definitions, ensuring the spectral purity of the agricultural baseline.

However, global agricultural systems are highly heterogeneous. To address the spatial sparsity of the Core set in specific complex environments, we generated the spatially augmented sample set (Fig. 1b). These samples were collected using a global equal area sampling strategy, explicitly targeted to fill spatiotemporal gaps in under-represented ecoregions. This densification ensures that the spectral diversity of specific cropping systems—such as smallholder mosaics in Sub-Saharan Africa or extensive soy monocultures in South America—is adequately represented. This process yielded an additional 332,471 sample locations.

2.3.2 Independent validation samples

To guarantee an objective assessment of the GACED30 product, we employed a strictly independent validation set, isolated entirely from the model calibration process. This subset was extracted from the stable cropland records of the FAST validation set (Li et al., 2017). Crucially, the FAST project generated this validation set using a sampling protocol entirely distinct from the training data: while training sites were manually selected for spectral representativeness, validation samples were allocated via a probabilistic global equal-area random sampling design. This probability-based framework ensures that the validation locations are spatially uncorrelated with the training samples and provide an unbiased representation of global land cover heterogeneity. Furthermore, unlike our training strategy, these validation samples were not subject to the GLAD-based stratification or any automated filtering, thereby preventing the propagation of prior-map biases into the accuracy assessment. The final validation set comprises 29,226 sample locations, including 3,268 confirmed stable cropland samples, providing a rigorous and independent benchmark.

2.4 National Agricultural Statistics (FAOSTAT)

To evaluate the agreement between our satellite-derived product and official national statistics, we compared the annual cropland area estimates from GACED30 with data from the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT) (FAO, 2022). As the centralized repository for national censuses and surveys reported by member countries, FAOSTAT constitutes the authoritative record for monitoring global food security and Sustainable Development Goals (SDGs). We accessed the land use statistic data (https://www.fao.org/faostat/en/#data/RL) to derive statistical baselines for 132 countries spanning the period from 2000 to 2024.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

3 Methods

The framework for generating the GACED30 dataset is illustrated in Fig. 2. This process integrates multi-source Earth observation data with advanced machine learning to produce a consistent 25-year record. The workflow comprises five main components: (1) feature engineering to construct a high-dimensional spectra-temporal space using SDC30; (2) annual cropland probability mapping utilizing spectral-semantic sample alignment and the CatBoost classifier; (3) spatio-temporal cropland extent refinement via temporal rotation logic and morphological field reconstruction; (4) derivation of Structural Evolution Indicators including cropping frequency and trend analysis; and (5) a comprehensive validation against independent samples and statistical datasets.

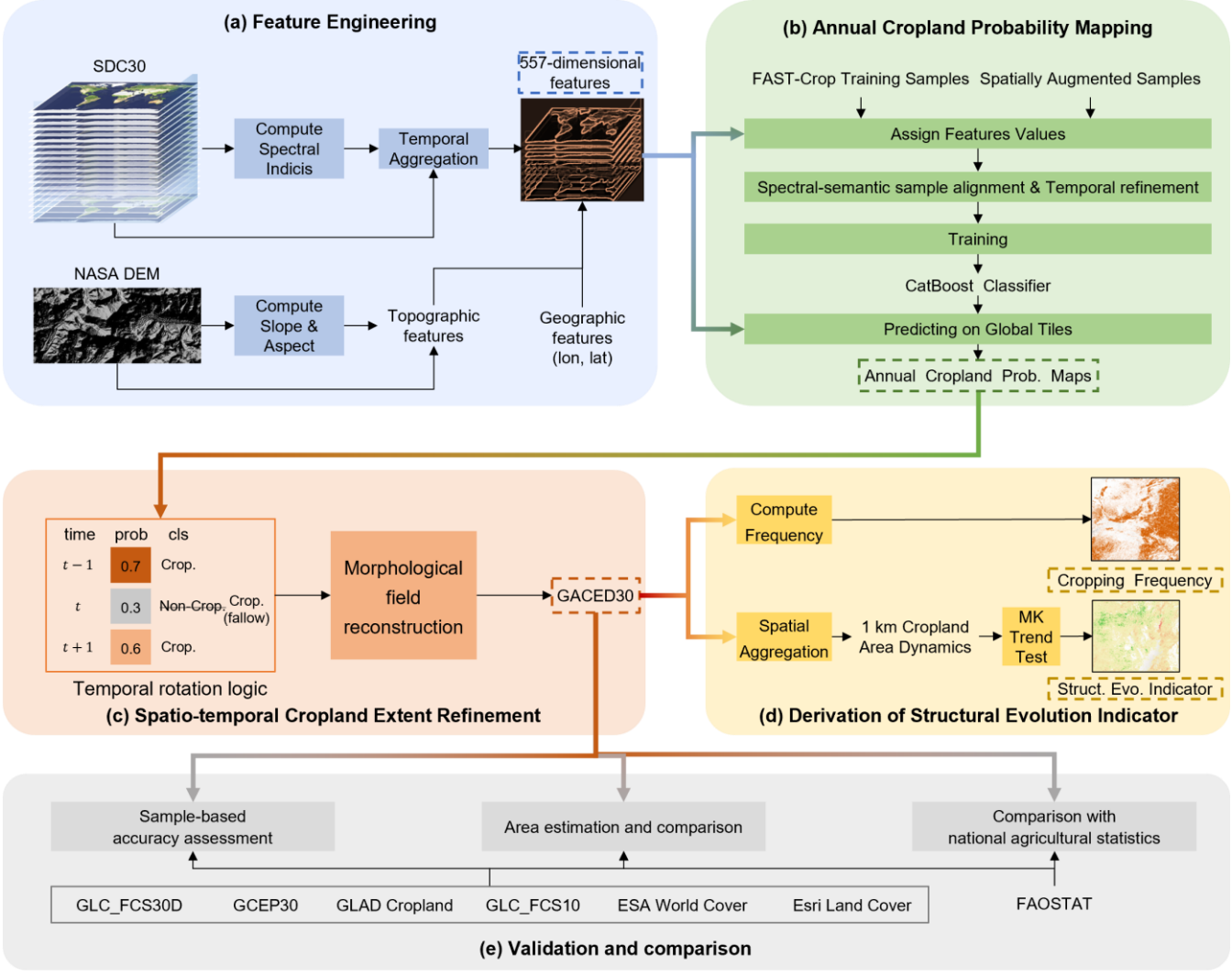


Figure 2. The workflow for generating the GACED30 dataset. (a) Feature engineering using SDC30 and NASADEM data. (b) Annual cropland probability mapping via spectral-semantic sample alignment and CatBoost classification. (c) Spatio-

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

temporal cropland extent refinement using temporal rotation logic and morphological field reconstruction. (d) Derivation of Structural Evolution Indicators including cropping frequency and trend analysis. (e) Sample-based accuracy assessment, global area estimation, and inter-comparison with products and statistics.

3.1 Feature engineering

To distinguish active cropland from natural vegetation, relying on static spectral signatures is insufficient. The defining characteristic of agriculture is its phenological trajectory—the rapid, human-induced cycle of sowing, biomass accumulation, and harvest. To capture this dynamic signal, we constructed a high-dimensional feature space based on the SDC30 time series (Fig. 2a).

While SDC30 offers daily resolution, global agricultural monitoring requires balancing temporal fidelity with computational feasibility. Therefore, we aggregated the daily reflectance into 8-day composites using a median reducer. This temporal cadence results in 46 uniformly distributed observation steps per year, a density that is sufficient to reconstruct complete growth cycles (including double-cropping systems) while smoothing out residual noise. For each of the 46 time steps, we extracted 12 spectral components optimized for agricultural monitoring:

- Basic Spectral Bands (6): Blue, Green, Red, NIR, SWIR1, and SWIR2
- Biomass and Structure Indices (3): NDVI (general greenness), kNDVI (to minimize saturation in high-biomass crops) (Camps-Valls et al., 2021), and MNDWI (to capture paddy rice inundation).
- Tasseled Cap Components (3): Brightness, Greenness, and Wetness (to enhance soil and moisture detection).

This process generated a dense time-series vector of 552 features per pixel.

To further constrain the model against false positives in non-arable terrain, we appended five auxiliary features to the spectral vector. We utilized NASADEM to derive Elevation, Slope, and Aspect, helping to distinguish flat arable land from natural vegetation on steep slopes (e.g., >25°) where mechanization is impossible. Additionally, pixel-level Latitude and Longitude were included not merely for location, but to implicitly encode agro-climatic zones, allowing the model to adapt to latitudinal variations in crop calendars (e.g., winter wheat in the north vs. spring wheat in the south).

The final input vector comprised 557 features per pixel, combining the phenological depth of the time series with the physical constraints of the terrain.

3.2 Annual cropland probability mapping

To generate continuous and consistent annual probability maps, we employed a mapping workflow (Fig. 2b) that first constructs a spatially extensive, thematically consistent training samples via spectral-semantic sample alignment and temporal refinement (Sect. 3.2.1), and subsequently utilizes year-specific CatBoost models to estimate pixel-wise cropland probabilities across the full 2000–2024 archive (Sect. 3.2.2).

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

3.2.1 Spectral-semantic sample alignment and temporal refinement

To construct a spatially representative and thematically consistent training library, we implemented a sample alignment strategy that harmonizes the high semantic fidelity of expert annotations with the global spatial coverage of existing products.

While the FAST-Crop sample set serves as the baseline, its spatial distribution is insufficient for global-scale supervised learning. To address this, we generated a candidate pool of approximately 360,000 locations using a Global Equal Area sampling design, stratified by the GLAD Cropland product (Potapov et al., 2022) to target under-represented ecoregions. However, direct utilization of these locations carries the risk of introducing systematic errors due to definition-level discrepancies between existing products and the GACED30 standard, particularly regarding the inclusion of permanent crops (orchards), agricultural structures (greenhouses), and the specific treatment of active fallow phases.

To strictly filter these definition-level conflicts, we implemented a spectral-semantic sample alignment protocol. We trained a preliminary CatBoost classifier using exclusively the expert-annotated FAST-Crop samples, thereby encapsulating the strict spectral-phenological definition of active cultivation. This expert-driven model was then applied to verify the candidate locations. We retained only those samples where the model prediction agreed with the spatial prior label, effectively filtering out ambiguous samples that satisfy generalized vegetation criteria but fail the specific phenological requirements of active cropping. This process yielded a refined spatially augmented set containing approximately 25,000 confirmed cropland and 307,000 non-cropland samples that inherit the high thematic fidelity of the FAST standards.

Finally, extending these static spatial samples across the full 2000–2024 study period relies on an assumption of temporal stability, which introduces label noise in pixels undergoing land use change (e.g., recent reclamation or urbanization). To resolve this, we integrated CleanLab, a data-centric AI framework based on Confident Learning (Northcutt et al., 2021). By estimating the joint distribution of noisy and true labels via cross-validation, CleanLab automatically identifies and prunes instances where the spectral signal of a specific year contradicts the assigned label. This automated cleaning ensures that the final training library is temporally dynamic, eliminating erroneous training points during their transition years while preserving valid samples during stable periods.

3.2.2 Annual probability estimation

Following the construction of the training sample set, we trained independent, year-specific classification models to account for inter-annual phenological variability. For each year from 2000 to 2024, the corresponding subset of valid training samples was extracted to train a specific CatBoost classifier (Prokhorenkova et al., 2018). We selected CatBoost as the inference engine for its superior generalization capability via ordered boosting, which mitigates the prediction shift often observed in standard gradient boosting decision trees. Furthermore, CatBoost utilizes oblivious decision trees to efficiently handle the high-dimensional nature of our feature space (557 spectra-temporal features) without the need for manual feature reduction, a capability successfully demonstrated in the ESA WorldCover global mapping project (Zanaga et al., 2022).

The trained year-specific models were then applied to the corresponding annual SDC30 feature stacks to generate cropland probability maps for the entire study period. This process produced a continuous 25-year record of pixel-wise

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

probability values, representing the likelihood of cropland extent. All training and inference computations were executed on the high-performance iEarth remote sensing cloud computing platform hosted by Pengcheng Laboratory.

3.3 Spatio-temporal cropland extent refinement

The raw probability maps generated in the previous step capture annual phenological signals but inevitably contain "salt-and-pepper" noise and temporal flickering caused by cloud contamination or phenological ambiguity. A particular challenge lies in the spectral confusion between "active fallow" and "natural bareland"—both exhibit low vegetation indices, yet one is an integral phase of the cropping cycle (classified as cropland). To transform these pixel-level probabilities into a coherent, high-quality cropland dataset, we applied a spatiotemporal refinement framework designed to reconstruct the structural integrity and temporal continuity of agricultural systems (Fig. 2c).

3.3.1 Temporal rotation logic (fallow recovery)

To resolve the ambiguity between fallow phases and non-crop transitions, we implemented a simplified temporal consistency rule targeting pixels with indeterminate probabilities. Specifically, we identified pixels in the uncertainty interval (Probability $P \in [0.3, 0.5]$), which are typically classified as non-crop in standard binary mapping.

We applied a "Rotation Continuity Rule": if a pixel in year t falls within this uncertainty range but is confirmed as active cropland ($P > 0.5$) in both the preceding ($t-1$) and succeeding ($t+1$) years, it is reclassified as Active Fallow (i.e., Cropland). This straightforward logic effectively bridges temporary gaps in the rotation cycle caused by fallowing or weak phenological signals, without introducing false positives from permanent land use changes (e.g., urbanization) where probabilities typically drop below 0.3.

3.3.2 Morphological field reconstruction

Following temporal refinement, the probability maps were binarized to generate initial boolean masks. We then applied morphological spatial constraints to enforce agronomic structural plausibility. Croplands, particularly in zones of intensive agriculture, typically manifest as large, contiguous patches rather than isolated or scattered pixels. Raw pixel-based classification, however, often results in "salt-and-pepper" noise that misrepresents field boundaries. To address this, we designed a morphological filter to identify and remove spurious cropland detections that violate shape regularity. We specifically targeted and removed artifacts defined by weak connectivity or insufficient area, including:

- Isolated single pixels lacking immediate neighbours
- Small linear fragments (e.g., 1×2 pixel strips) that do not form a cohesive field structure
- Diagonally connected clusters (checkerboard patterns) that lack 4-connectivity

By filtering out these fragmented clusters, we improved the visual continuity and analytical coherence of the final GACED30 dataset, ensuring the maps reflect the topological reality of managed agricultural parcels.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

3.4 Derivation of Structural Evolution Indicators

To accurately quantify the spatiotemporal dynamics of global agricultural systems from 2000 to 2024, we developed a grid-based analytical framework that synthesizes management intensity with directional trends (Fig. 2d).

First, to characterize the stability of cultivation, we calculated the pixel-level Cropping Frequency (F_{crop}), defined as the proportion of years a pixel was identified as active cropland over the 25-year period. This metric serves as a proxy for management intensity, effectively distinguishing "Core Croplands" (high frequency, e.g., >80\%, indicating continuous cultivation) from "Marginal Zones" (low-to-intermediate frequency). While low-frequency areas typically represent shifting cultivation or frontier zones, frequency alone cannot determine the trajectory of change—for example, a frequency of 20% could result from recent reclamation (expansion) or historical abandonment (reduction).

To resolve this ambiguity and identify the directional evolution of these landscapes, we derived Structural Evolution Indicators using a robust trend analysis framework based on the GACED30. To ensure the geometric accuracy of these calculations at the global scale, we adopted the practical principles outlined by Tyukavina et al. (2025), utilizing a Global Equal Area projection to eliminate area distortion inherent in geographic coordinate systems during pixel summation.

To mitigate the "salt-and-pepper" noise associated with high-resolution annual classification and to capture robust landscape-level changes, we aggregated the 30 m binary cropland maps into 1×1 km grid cells. For each grid cell, we calculated the annual cropland area in hectares, resulting in a continuous 25-year time series. We applied the non-parametric Mann-Kendall (MK) trend test combined with the Theil-Sen estimator to this time series. This approach provides a rigorous statistical basis for identifying change, yielding three key metrics for each grid: the trend slope (β, unit: ha/year), the statistical significance (p-value), and the confidence intervals (upper and lower bounds of the slope). Based on these metrics, we categorized the global cropland dynamics into three distinct trajectories:

- Expansion: Areas with a statistically significant positive trend ($\mathrm{p} < 0.01$) and a slope $\beta > 0.2$.
- Reduction: Areas with a statistically significant negative trend ($\mathrm{p} < 0.01$) and a slope $\beta < -0.2$.
- Stable: Areas where the trend was not statistically significant ($\mathrm{p} < 0.01$) or the magnitude of change was negligible ($-0.2 \leq \beta \leq 0.2$).

The threshold of 0.2 ha/year (representing a 0.2% annual crop change within a 100 ha grid) was selected to filter out minor fluctuations caused by temporary fallowing or inter-annual classification instability, ensuring that only substantial land use conversions are reported.

This continuous time-series methodology offers significant advantages over the epoch-based comparison methods used in existing global datasets, such as GLAD Cropland (Potapov et al., 2022). The GLAD approach derives change by comparing composite metrics between multi-year intervals (e.g., 2000–2003 vs. 2016–2019). While effective for stabilizing noisy data, epoch-based methods are sensitive to the selection of the baseline and end years; an extreme weather event or temporary crop rotation during the epoch windows can bias the detection of "Gross Gain" or "Gross Loss".

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

In contrast, our approach utilizes the full annual density of the GACED30 (2000–2024). By evaluating the monotonic trend across 25 data points rather than two static snapshots, our method is robust against short-term anomalies and provides a more stable characterization of long-term dynamics. Furthermore, the inclusion of statistical significance tests (p-value) and confidence intervals for the trend slopes allows us to distinguish between genuine land cover change and random variability with a quantifiable level of certainty, providing a more reliable evidence base for assessing the acceleration of global agricultural expansion.

3.5 Validation and comparison

To comprehensively assess the quality and reliability of the GACED30 dataset, we adopted a multi-faceted validation strategy involving three distinct approaches (Fig. 2e): (1) quantitative accuracy assessment using FAST-Crop validation sample sets; (2) global area estimation and inter-comparison with existing state-of-the-art global land cover and cropland products; and (3) statistical correlation analysis with national agricultural statistics from FAOSTAT.

3.5.1 Sample-based accuracy assessment

We evaluated the pixel-level accuracy of GACED30 using the independent FAST-Crop validation set. To ensure a consistent evaluation across different satellite products with varying availability, we employed a nearest neighbor temporal matching strategy. For any given validation sample labeled in year T_{sample}, we selected the map layer from the product (GACED30 or reference products like GLAD Cropland) with the year T_{product} that minimized the absolute temporal difference $|T_{\text{sample}} - T_{\text{product}}|$. This ensures that the validation reflects the product's performance closest to the ground reference date.

We calculated a binary confusion matrix (Cropland vs. Non-Cropland) to derive standard quantitative accuracy metrics. Let n_{ij} denote the number of samples classified as class i in the map but belonging to class j in the reference data, where $k = 2$ is the number of classes, and N is the total number of samples. The specific metrics are calculated as follows.

Overall Accuracy (OA) represents the proportion of correctly classified pixels:

$$\text{OA} = \frac{\sum_{i=1}^{k} n_{ii}}{N} \times 100$$

Producer's Accuracy (PA), which relates to omission error, indicates the probability that a ground reference pixel is correctly classified in the map:

$$PA_i = \frac{n_{ii}}{n_{+i}} \times 100$$

where n_{+i} is the total number of reference samples for class i.

User's Accuracy (UA), which relates to commission error, indicates the probability that a pixel classified as class i on the map actually represents that class on the ground:

$$UA_i = \frac{n_{ii}}{n_{i+}} \times 100$$

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

where n_{i+} is the total number of map predictions for class i.

The Kappa Coefficient (κ) measures the agreement between the classification map and reference data while accounting for chance agreement:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the observed agreement ($\mathrm{OA}/100$), and P_e is the hypothetical probability of chance agreement, calculated as:

$$P_e = \frac{\sum_{i=1}^{k}(n_{i+} \times n_{+i})}{N^2}$$

To benchmark the performance of GACED30, we applied these metrics to conduct a comparative accuracy assessment against the GLAD Cropland product (Potapov et al., 2022), as well as other available global land cover datasets including 30-m products (GCEP30 (Thenkabail et al., 2021), GLC_FCS30D (Zhang et al., 2023)) and 10-m products (GLC_FCS10 (Zhang et al., 2025c), ESA World Cover (Zanaga et al., 2022), Esri Land Cover (Karra et al., 2021)), ensuring all products were evaluated against the identical FAST-Crop reference baseline.

3.5.2 Global area estimation and inter-comparison

To ensure accurate area estimation within a consistent and reproducible framework, we utilized Open Data Cube (ODC) technology to manage and process all multi-source datasets (Killough, 2018; Killough et al., 2020). Rather than reprojecting all data into a single global system—which can introduce resampling artifacts—we maintained the native projections of each product and calculated areas using projection-specific protocols derived from Tyukavina et al. (2025).

For datasets provided in UTM projections (including GACED30 and Esri Land Cover), we calculated the area separately for each UTM zone to minimize geometric distortion, summing the results to derive global totals as per the zonal guidelines in Tyukavina et al. (2025). For datasets provided in EPSG:4326 (Geographic Lat/Lon) (including ESA WorldCover, GLAD Cropland, GLC_FCS30D, and the geographic version of Esri Land Cover), we retained the native grid and calculated pixel areas using the latitude-weighted geodetic formula described by Tyukavina et al. (2025), which accounts for the convergence of meridians towards the poles.

We applied sample-based area estimation exclusively to the GACED30 dataset. This restriction was necessary because the FAST-Crop validation sample set was developed using a classification system and definition strictly aligned with GACED30; applying these samples to correct other products with differing definitions (e.g., ESA WorldCover's exclusion of orchards) would introduce significant bias rather than correct it. For GACED30, we integrated the map data with the confusion matrix derived from the FAST-Crop validation samples to calculate the "unbiased estimated area" and its associated uncertainty. The Standard Error of the area estimate was calculated following the good practice protocols of Olofsson et al. (2014) and Tyukavina et al. (2025), utilizing the stratum weights (mapped class proportions) and the variance of the sample proportions within each stratum to construct 95% confidence intervals for the annual cropland area.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

3.5.3 Comparison with national agricultural statistics

National statistics provide a crucial independent benchmark for evaluating the reliability of satellite-derived area estimates. We compared the aggregated national cropland areas from GACED30 with official statistics from FAOSTAT.

To facilitate a rigorous comparison, we reconciled the FAOSTAT aggregates to match the definition of GACED30 (Sect. 2.1). The standard FAO definition of "Cropland" (Item 6620) comprises "Arable land" (Item 6621) and "Permanent crops" (Item 6650). However, the sub-category "Arable land" includes "Temporary meadows and pastures" (Item 6633)—defined as land cultivated with herbaceous forage for less than five years. As GACED30 and the majority of high-resolution cropland products (Table 1) explicitly exclude natural and semi-natural grasslands, we calculated the reference cropland area by subtracting "Temporary meadows and pastures" (Item 6633) from the total "Cropland" aggregate (Item 6620), which is also suggested by Tubiello et al. (2023b). This adjusted metric, encompassing annual temporary crops, temporary fallow, and permanent woody crops (orchards), aligns with our mapped classes and serves as the benchmark for assessing the accuracy and trends of the GACED30 dataset. We utilized the Coefficient of Determination (R^2) and Normalized Root Mean Square Error (NRMSE) to quantify the correlation between the GACED30-derived areas and the adjusted FAOSTAT statistics across all available countries.

Beyond static area comparisons, we further evaluated the consistency of long-term cropland dynamics by benchmarking the net area change derived from GACED30 against FAOSTAT trends. We calculated the net area change for each country over the study period (2000–2024). To mitigate the impact of inter-annual fluctuations (e.g., extreme weather events) and reporting inconsistencies, we adopted a robust multi-year averaging approach, defining net change as the difference between the mean area of the last three available years and the mean area of the first three available years. We restricted this analysis to countries with a temporal record spanning at least 10 years and reporting "significant" land use dynamics (defined as a net change > 0.5% of the total cropland area or > 1,000 ha) in FAO statistics. This filtering ensures that the comparison focuses on genuine structural trends rather than minor statistical noise. The agreement was quantified using two metrics: the Direction Match Rate, which measures the percentage of countries where GACED30 and FAOSTAT report the same trajectory (expansion or contraction), and the Weighted Match Rate, which weights this agreement by the total cropland area of each country to emphasize performance in major agricultural nations.

4 Results

4.1 Quantitative evaluation

4.1.1 Accuracy assessment

To evaluate the reliability of the GACED30 dataset, we conducted a rigorous pixel-level accuracy assessment using the FAST-Crop validation sample set. This dataset, derived from expert interpretation of multi-seasonal imagery, serves as a high-quality, independent benchmark strictly excluded from the training process. We benchmarked GACED30 against six

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

state-of-the-art global 30-m and 10-m land cover and cropland products: GLAD Cropland, GCEP30, GLC_FCS30D, GLC_FCS10, ESA WorldCover and Esri Land Cover.

As detailed in Table 2, GACED30 achieved the highest performance across all accuracy metrics, demonstrating its robustness in capturing global cropland dynamics. GACED30 achieved an Overall Accuracy (OA) of 96.5% and a Kappa coefficient of 0.825, significantly outperforming the widely used GLAD Cropland product (OA: 94.6%, Kappa: 0.714), 10-m ESA World Cover (OA: 94.1%, Kappa: 0.686) and long-term GLC_FCS30D (OA: 90.1%, Kappa: 0.825). Crucially, the F1 score—which balances precision and recall and is less sensitive to class imbalance—reveals a substantial advantage for GACED30 (0.844). This score is markedly higher than those of GLAD Cropland (0.744), ESA WorldCover (0.722), and Esri Land Cover (0.689).

Table 2: Comparison of mapping accuracy

Product	F_1	OA	UA	PA	Kappa
GACED30	0.844	0.965	0.883	0.808	0.825
GLAD Cropland	0.744	0.946	0.835	0.671	0.714
GCEP30	0.597	0.898	0.550	0.652	0.539
GLC_FCS30D	0.626	0.901	0.552	0.722	0.825
GLC_FCS10	0.657	0.917	0.633	0.683	0.610
ESA World Cover	0.722	0.941	0.790	0.664	0.689
Esri Land Cover	0.689	0.931	0.727	0.655	0.651

It is important to note that the lower metrics of peer products partly stem from systematic definitional discrepancies (e.g., the treatment of pastures and woody crops) and temporal mismatches. However, GACED30's superior performance against the strict, FAO-aligned validation set objectively confirms its targeted agronomic fidelity. This result validates the effectiveness of our phenology-adaptive framework, demonstrating that integrating spectral-semantic sample alignment successfully resolves the definitional ambiguities and temporal inconsistencies.

4.1.2 Statistical consistency with FAO national statistics

To evaluate the reliability of GACED30 at the national scale, we aggregated pixel-level classifications into national totals and compared them with official statistics from FAOSTAT. GACED30 estimates demonstrate a robust statistical agreement with reported data across 132 countries, achieving an R^2 of 0.95 and a Normalized Root Mean Square Error (NRMSE) of 3.7% (Fig. 3a). These metrics indicate that GACED30 captures the inter-country variability of agricultural production capacity with high fidelity.

The performance of GACED30 is particularly significant when contextualized against the comprehensive validation of global cropland products conducted by Tubiello et al. (2023a). In their assessment of existing 30-meter global products (Table D1), established datasets such as GLC_FCS30 achieved a lower R^2 of 0.89 and a significantly higher NRMSE of 6.5%, while GlobeLand30 achieved an R^2 of 0.94 with an NRMSE of 6.1%. Our product aligns closely with the accuracy of

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

the 10-meter ESA WorldCover product (R^2 0.95, NRMSE 3.3%) and the meta-analysis CAM aggregation dataset (R^2 0.95, NRMSE 3.0%). This comparison demonstrates that despite being an automated annual product derived from continuous change detection, GACED30 maintains a statistical precision superior to many existing 30-meter dynamic products and comparable to static, visually refined maps.

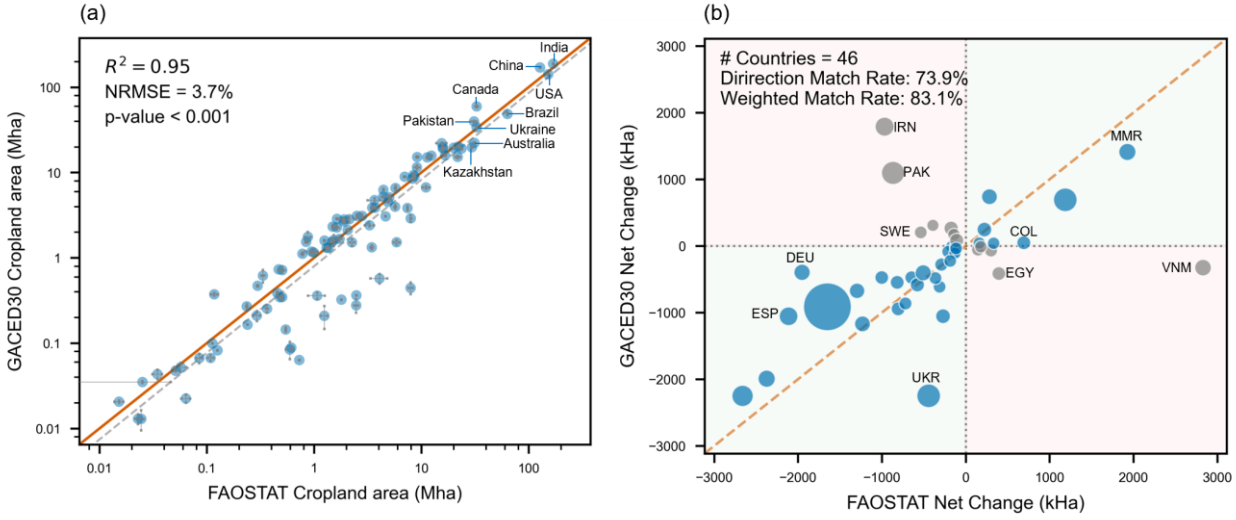


Figure 3. Consistency between GACED30 and FAOSTAT national statistics. (a) Comparison of mean annual cropland area (Mha) for 132 countries. The solid orange line indicates the 1:1 relationship; error bars represent the inter-annual standard deviation. (b) Comparison of long-term net area change (kHa) from 2000 to 2024 for 42 countries with significant reported dynamics. Bubbles are scaled by total cropland area, with blue indicating directional agreement and grey indicating disagreement.

Beyond static area totals, we further verified the product's ability to reproduce the temporal trajectory of agricultural change. We compared the long-term net trend (2000–2024) derived from GACED30 against FAOSTAT for 42 nations reporting significant land cover dynamics (Fig. 3b). GACED30 achieved a directional match rate of 73.9%, indicating that the product correctly identifies whether a country's cropland area is expanding or contracting in the vast majority of cases. When weighted by national cropland area, the consistency remains robust at 83.1%. This confirms that GACED30 accurately captures the structural trends of major agricultural producers despite the differing temporal granularities of satellite observations versus census reporting.

While the overall correlation between GACED30 and FAOSTAT is robust, observable discrepancies persist, primarily driven by the 'mixed pixel' effect inherent to 30-meter resolution analysis (Pax-Lenney and Woodcock, 1997). In regions dominated by smallholder farming or highly fragmented agricultural mosaics, the physical dimensions of cultivated plots are often smaller than the satellite pixel footprint. Consequently, these sub-pixel crop areas are frequently misclassified when the spectral signature is dominated by surrounding non-crop vegetation. This spatial resolution constraint creates an inevitable

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

divergence from FAOSTAT census data, which aggregates administrative records of land holdings independent of their geometric detectability.

Conversely, for China, GACED30 estimates are higher than the standard FAOSTAT reporting. This discrepancy is a consistent phenomenon observed in recent remote sensing literature (Yu et al., 2021). Official statistics for China have historically under-reported cultivated land due to complex land tenure and reporting systems, whereas satellite-based estimates consistently identify larger extents of agricultural activity. The higher estimate in GACED30 aligns with these independent satellite observations, suggesting that our product effectively captures fragmented smallholder landscapes that may be omitted or aggregated in administrative records.

Furthermore, the error bars in Fig. 3 provide critical insights into the temporal dynamics of cultivation. Countries with large vertical error bars, such as Australia, exhibit high inter-annual variability in the GACED30 dataset. This accurately reflects the ecological reality of large-scale rainfed agriculture in these regions, where vast areas may be left fallow during drought years or periods of low economic return (Xie et al., 2024). In contrast, the significantly smaller horizontal error bars associated with FAO data indicate that static statistical reporting often smooths out these dynamic fluctuations. This demonstrates that GACED30 captures the physical reality of inter-annual land use intensity and fallow cycles more effectively than traditional statistical reporting methods.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Open Access Earth System Science Data Discussions

4.2 Qualitative evaluation

4.2.1 Inter-comparison in challenging landscapes

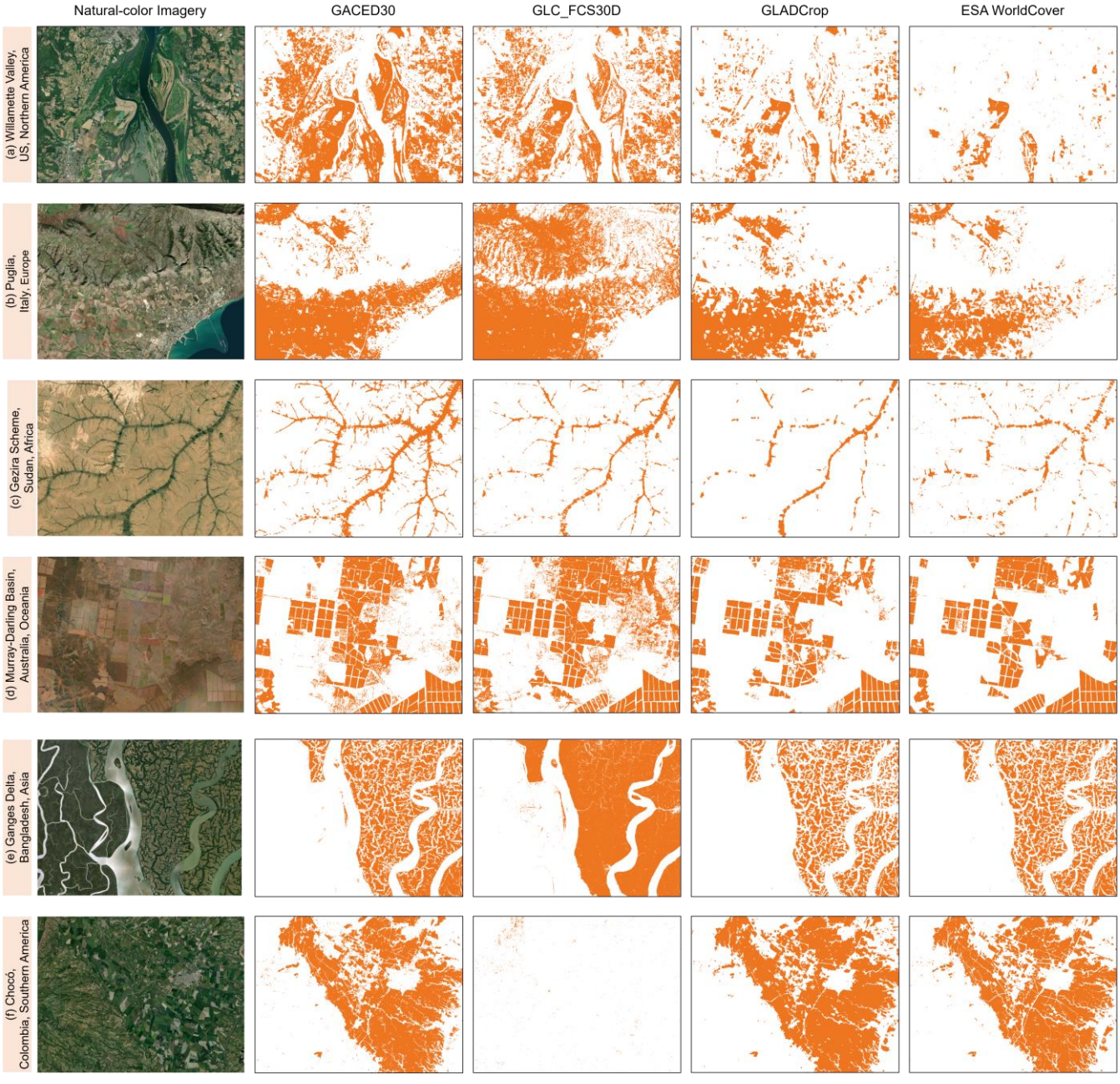


Figure 4. Visual comparison of GACED30 with GLC_FCS30D, GLAD Cropland, and ESA WorldCover across six representative challenging agricultural landscapes. The panels illustrate performance in: (a) Willamette Valley, USA; (b) Puglia, Italy; (c) Gezira Scheme, Sudan; (d) Murray-Darling Basin, Australia; (e) Ganges Delta, Bangladesh; and (f) Chocó

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Department, Colombia. Background satellite imagery is sourced from Esri World Imagery Wayback (June 2020) (Sources: Esri | Powered by Esri).

To demonstrate the robustness of GACED30, we selected representative case studies from challenging agricultural landscapes across all continents, purposely moving beyond typical breadbaskets to test performance in spectrally complex and data-sparse regions. We conducted a visual inter-comparison using 2020 as the baseline year to align with the availability of high-quality validation products such as ESA WorldCover and GLC_FCS30D, while utilizing the 2019 epoch for GLAD Cropland.

In North America, the Willamette Valley in the USA (Fig. 4a) presents a unique challenge as a major global hub for grass seed production, where managed crops are spectrally almost identical to pasture and hay (Mueller-Warrant et al., 2011). While GLC_FCS30D and ESA WorldCover often misclassify these managed grass crops as generic grassland, GACED30 successfully distinguishes the active cultivation cycle of seed grass from permanent pastures. Similarly, in the drought-susceptible Murray-Darling Basin of Australia (Fig. 4d), rainfed fields often lie fallow for extended periods, creating spectral confusion between fallow land and semi-arid scrub (Xie et al., 2024). In this region, GLC_FCS30D tends to overestimate cropland extent by including natural scrub, whereas GACED30's statistical trend analysis effectively filters out these fluctuations, separating genuine abandonment from the rotational fallow characteristic of viable farms.

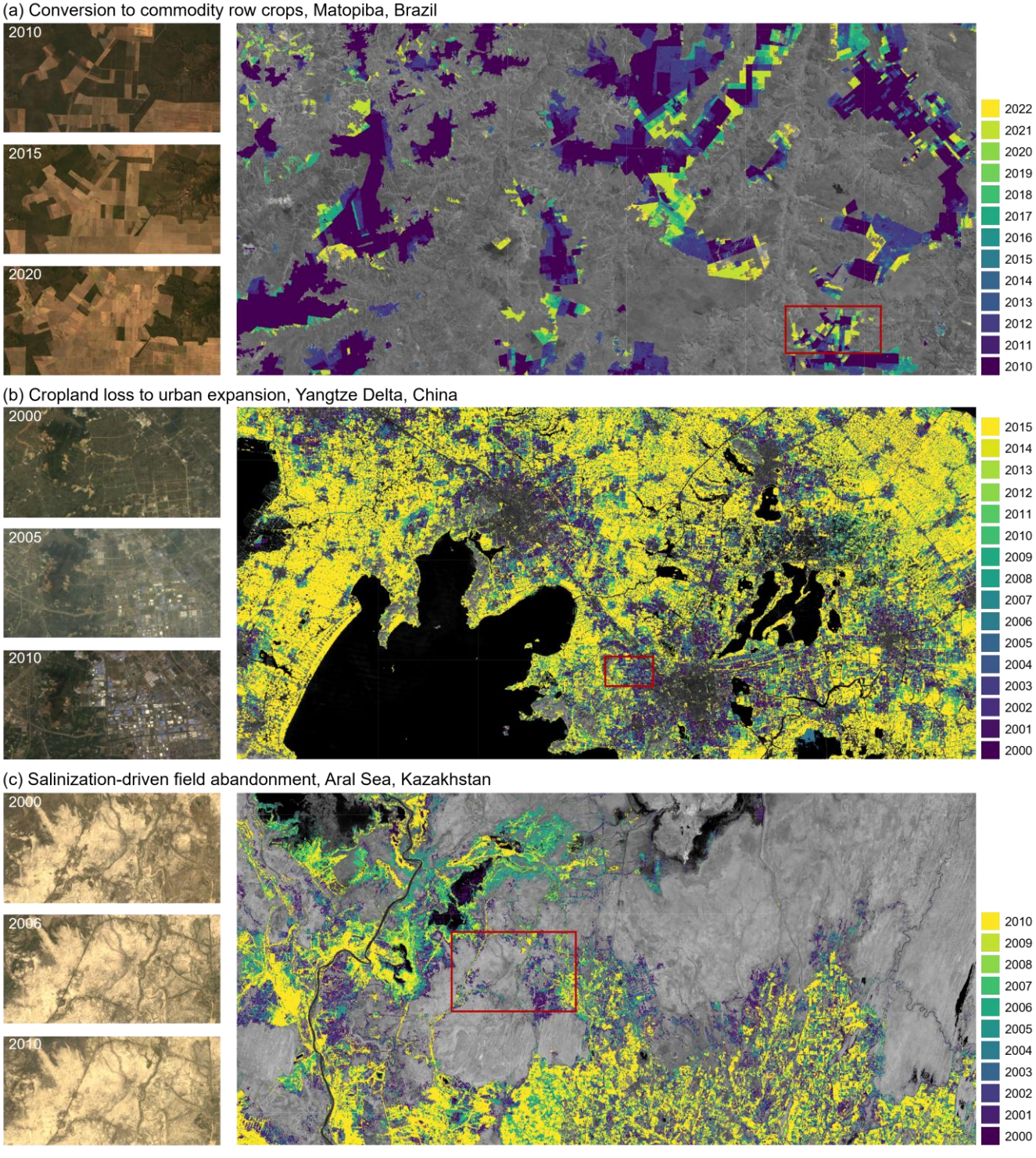
Moving to Europe, the landscape of Puglia, Italy (Fig. 4b) is characterized by ancient olive groves intermixed with winter wheat, creating a complex mixed-pixel environment (Damianidis et al., 2021). Unlike GLAD Cropland, which explicitly excludes orchards and consequently omits these groves, GACED30 adopts a comprehensive definition that accurately captures this mixed woody-herbaceous agro-ecosystem. In the arid Gezira Scheme of Sudan (Fig. 4c), the distinction between fallow fields and the surrounding bare desert is critical. While epoch-based products like GLAD often produce "checkerboard" artifacts by omitting plots that were fallow during the specific epoch window, GACED30 utilizes a 25-year cropping frequency to correctly identify these temporarily bare areas as part of the long-term irrigation rotation.

In the complex hydrological environment of the Ganges Delta, Bangladesh (Fig. 4e), distinguishing seasonal rice paddies from permanent aquaculture is difficult for single-snapshot maps due to the high fragmentation and small plot sizes. GACED30 effectively separates the seasonal phenological signal of rice from stable water bodies, reducing the commission errors seen in products like ESA WorldCover, which occasionally confuses flooded paddies with permanent wetlands. Finally, in the Chocó Department of Colombia (Fig. 4f), persistent cloud cover historically hinders optical mapping (Yang et al., 2025). By leveraging the gap-free reconstruction of the SDC30, GACED30 provides a spatially continuous map of subsistence agriculture in clearings that remain invisible or are marked as "no data" in standard optical products.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Open Access Earth System Science Data Discussions

4.2.2 Characterization of annual cropland dynamics



https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Figure 5. Spatiotemporal characterization of typical land use change events detected by GACED30 (2000–2022). The varying colors represent the specific year in which the land cover transition was first detected, illustrating the dataset's capacity to resolve the precise timing of dynamic events. (a) Agricultural Expansion: The progressive conversion of natural vegetation to commodity row crops in the Matopiba region, Brazil. (b) Cropland Loss: The irreversible displacement of peri-urban cropland by rapid urban expansion in the Yangtze River Delta, China. (c) Field Abandonment: The retreat of cultivation due to salinization and water stress in the Aral Sea basin, Kazakhstan.

While the spatial inter-comparison in Sect. 4.2.1 demonstrates the geometric accuracy of GACED30, the dataset's primary innovation lies in its temporal granularity. By leveraging the continuous annual record of the SDC30, GACED30 moves beyond the multi-year epochal aggregation used by products such as GLAD Cropland, enabling the precise temporal characterization of land use change events. Fig. 5 visualizes this capability across three distinct driver-response scenarios, where the pixel color denotes the specific year of land cover transition.

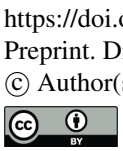
In the "Matopiba" region of Brazil (Fig. 5a), a global hotspot for agricultural intensification, GACED30 captures the rapid, contagious expansion of commodity row crops. Unlike epoch-based assessments that might flag a pixel as "converted" merely within a four-year window (e.g., 2011–2014), GACED30 resolves the specific year of conversion for individual fields. This level of detail reveals the spatial progression of the agricultural frontier, showing how cultivation spreads from established logistics corridors into the Cerrado biome year by year. This precise dating is critical for attributing land use change to specific annual market drivers or policy shifts.

Conversely, in the Yangtze River Delta of China (Fig. 5b), the dataset documents the irreversible loss of high-quality cropland to urbanization. The timeline of displacement (indicated by the spectral gradient from purple to yellow) aligns with the region's rapid economic development phases. GACED30 effectively identifies the exact year that fragmentation turns into permanent sealing, offering a tool for monitoring compliance with urban growth boundaries that coarser temporal products cannot provide.

Finally, in the Aral Sea region of Kazakhstan (Fig. 5c), GACED30 identifies the retreat of agriculture driven by environmental stress. Here, the signal is not of rapid conversion but of permanent abandonment due to soil salinization and water scarcity. By tracking the cessation of cropping activities annually, GACED30 distinguishes this terminal abandonment from temporary fallowing cycles. This capability to pinpoint the "year of exit" allows researchers to correlate agricultural collapse with specific hydrological drought events in the Syr Darya basin, a correlation that is obscured in datasets that aggregate observations over multiple years.

4.3 Spatiotemporal dynamics and structural evolution (2000–2024)

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026



4.3.1 Trends in global and regional cropland area

Based on the GACED30 time series, we estimate the global cropland area—defined as active annual and permanent cultivation excluding temporary pastures—to be approximately 1458.5 Mha in 2000, remaining relatively stable at 1483.3 Mha in 2020, and reaching 1488.5 Mha by 2024 (Fig. 6a).

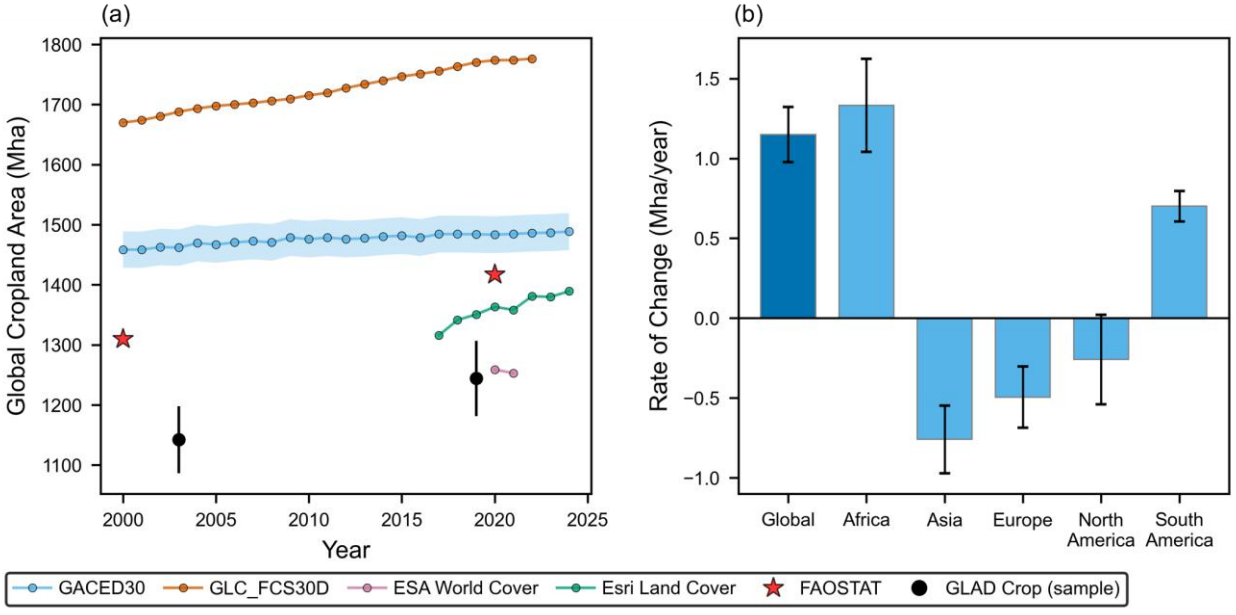


Figure 6. Inter-annual variations and trends of global cropland area from 2000 to 2024. (a) Global cropland area trajectories comparing GACED30 with GLC_FCS30D, Esri Land Cover, ESA WorldCover, and FAOSTAT benchmarks. (b) Global and continent-level trend analysis of crop extent area.

To validate the absolute magnitude of these estimates, we benchmarked GACED30 against FAOSTAT statistics, adopting an objective approach that harmonizes definitions to ensure comparability. The FAO "Cropland" aggregate (approx. 1562 Mha in 2020) comprises "Arable land" and "Permanent crops" (FAO, 2022) (Tubiello et al., 2023b). However, "Arable land" administratively includes "Temporary meadows and pastures" (approx. 145 Mha)—areas often characterized by natural forage rather than active cultivation. By adjusting the FAO statistics to exclude these pastures while retaining permanent crops (orchards, vineyards, and plantations), we established a reference benchmark of 1417 Mha for 2020. Our GACED30 estimate (1483.3 Mha) aligns remarkably well with this adjusted benchmark, exceeding it by only 4.7% (Fig. 6a). This close convergence suggests that GACED30 successfully captures the full spectrum of active agriculture, including the permanent woody crops often omitted by other remote sensing products.

The "Permanent Crop Gap" becomes evident when comparing GACED30 with other global datasets, as the definitional differences detailed in Table 1 directly drive the divergences observed in Fig. 6a. Products such as ESA WorldCover (1258.9 Mha in 2020) (Zanaga et al., 2022) and the sample-based GLAD Cropland (1244 Mha in 2019) (Potapov et al., 2022) report

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

significantly lower areas—roughly 200–240 Mha below the GACED30 estimate. As explicitly stated in their definitions (Table 1), these products strictly exclude "perennial woody crops" (e.g., fruit trees, coffee, rubber) to focus primarily on annual herbaceous crops. Similarly, Esri Land Cover (1363.2 Mha in 2020) (Karra et al., 2021) likely follows a similar exclusion logic, as its lower estimate suggests complex orchard systems are classified as forest or shrubland. In sharp contrast, GLC_FCS30D reports a significantly higher area (1773.6 Mha in 2020) (Zhang et al., 2023). This overestimation stems from a broader definition that includes not only "herbaceous cover cropland" and "tree or shrub cover cropland" but also likely conflates temporary pastures and rangelands with active agriculture (Table 1). GACED30 occupies a unique position by bridging this gap: it includes the permanent woody crops missed by GLAD and ESA WorldCover, while effectively filtering out the natural vegetation and pastures that inflate GLC_FCS30D estimates.

Regarding temporal dynamics, GACED30 reveals a conservative global expansion trend, with a net increase of approximately 30 Mha over the 24-year period (Fig. 6a). This finding echoes the conclusions of Tubiello et al. (2023b), who noted a discrepancy between FAO reports of relative global stability and satellite products showing massive expansion. While products like Esri Land Cover indicate rapid, linear growth (increasing by ~73 Mha from 2017 to 2024), GACED30's trajectory is far more stable, aligning with the FAO's observation of plateauing arable land in the Global North.

Regional analysis (Fig. 6b) further clarifies this trend. Statistically significant expansion is concentrated in the Global South (South America and Africa), driven by commodity crop expansion and smallholder extensification (Song et al., 2021). Conversely, the Global North (North America and Europe) exhibits widespread stability. By rigorously separating permanent cropland change from temporary variability, GACED30 confirms that while specific frontiers are expanding, the aggregate global cropland area has remained relatively stable, avoiding the inflation of trends caused by inter-annual spectral variability.

4.3.2 Patterns of management intensity and cropping frequency

Beyond the static delineation of cropland extent, the annual continuity of GACED30 allows for the derivation of "Cropping Frequency" (Fig. 7), a metric that quantifies the temporal stability of agricultural land cover over the 25-year study period. Unlike epoch-based datasets that aggregate multiple years to overcome data gaps, GACED30's gap-free annual record enables a precise calculation of how many years a pixel was classified as cropland between 2000 and 2024. It is important to clarify that in this context, frequency reflects the persistence of the "cropland" class—which encompasses both active cultivation and temporary fallow—rather than a direct measure of harvest intensity (e.g., double or triple cropping). Consequently, this metric effectively translates the binary "cropland/non-cropland" classification into a continuous gradient of stability, distinguishing between permanent agricultural zones and areas subject to intermittent land use or transition.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

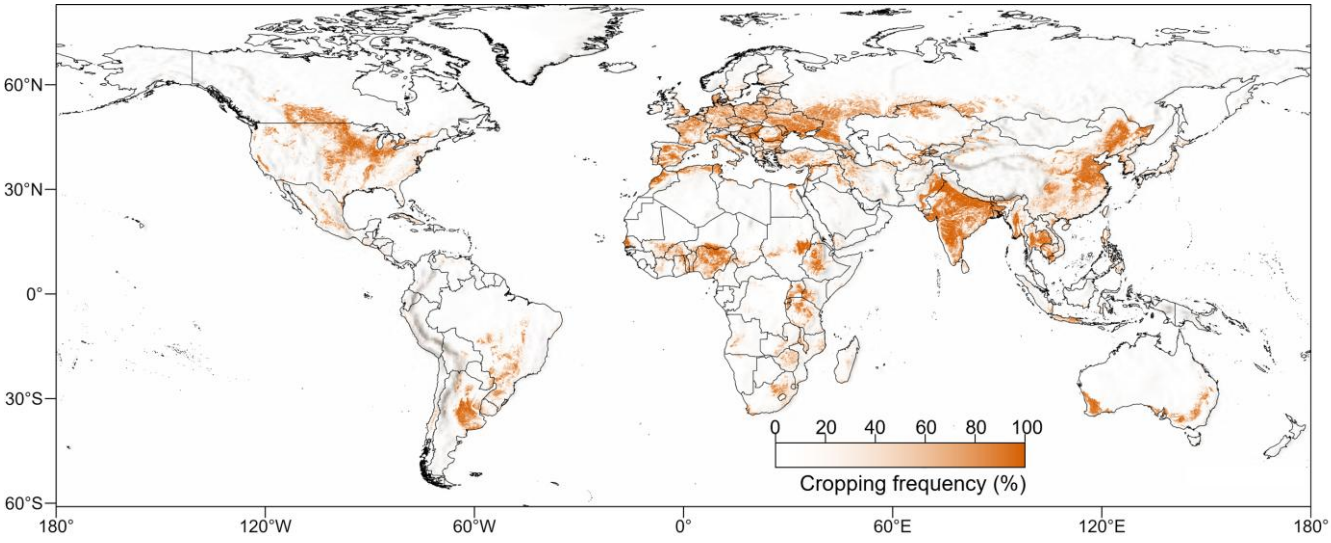


Figure 7. Global spatial distribution of cropping frequency derived from the GACED30 dataset (2000–2024). The pixel values represent the ratio of years classified as cropland to the total 25-year period, ranging from 0% (never cropped) to 100% (continuously cropped). The national administrative boundaries used in this map are sourced from the Resources and Environmental Science Data Platform.

Regions exhibiting a cropping frequency approaching 100% (represented by the darkest orange tones in Fig. 7) delineate the world's zones of stable, permanent agriculture. These align precisely with the primary global "breadbaskets," including the US Corn Belt, the North China Plain, and the Indo-Gangetic Plain. In these core production areas, the high consistency of the GACED30 signal validates the dataset's stability, confirming that the method successfully filters out the temporal noise (e.g., cloud contamination or phenological variability) that often causes spurious "flickering" in single-year products. Conversely, large areas within rainfed agricultural zones—specifically in the Sahel belt of Africa, Central Asia, and parts of Australia—display intermediate cropping frequencies (typically 40–70%). These intermediate values capture landscapes characterized by rotational systems or shifting cultivation, where land historically alternates between active status and long-term natural regeneration, or where the spectral signal of fallow land is highly variable.

Furthermore, low cropping frequency values (<30%) serve as a temporal signature for dynamic land cover transitions rather than stable low-intensity farming. The geographic distribution of these low-frequency pixels often reveals the footprint of recent change. For instance, in South America (specifically the Brazilian Cerrado) and Sub-Saharan Africa, low-frequency pixels concentrated at the peripheries of dense agricultural clusters typically represent frontier expansion—lands that were natural vegetation for the majority of the time series and have only recently been converted. In contrast, in Eastern Europe and the Russian Federation, similar low-frequency values often indicate historical croplands that were active only in the early 2000s before being abandoned. By quantifying this temporal persistence, GACED30 moves beyond static area estimation to highlight these "Signatures of Expansion and Retreat," setting the stage for the detailed quantification of these global reclamation and abandonment trends in the following section.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Open Access Earth System Science Data Discussions

615 **4.3.3 Spatiotemporal patterns of structural evolution**

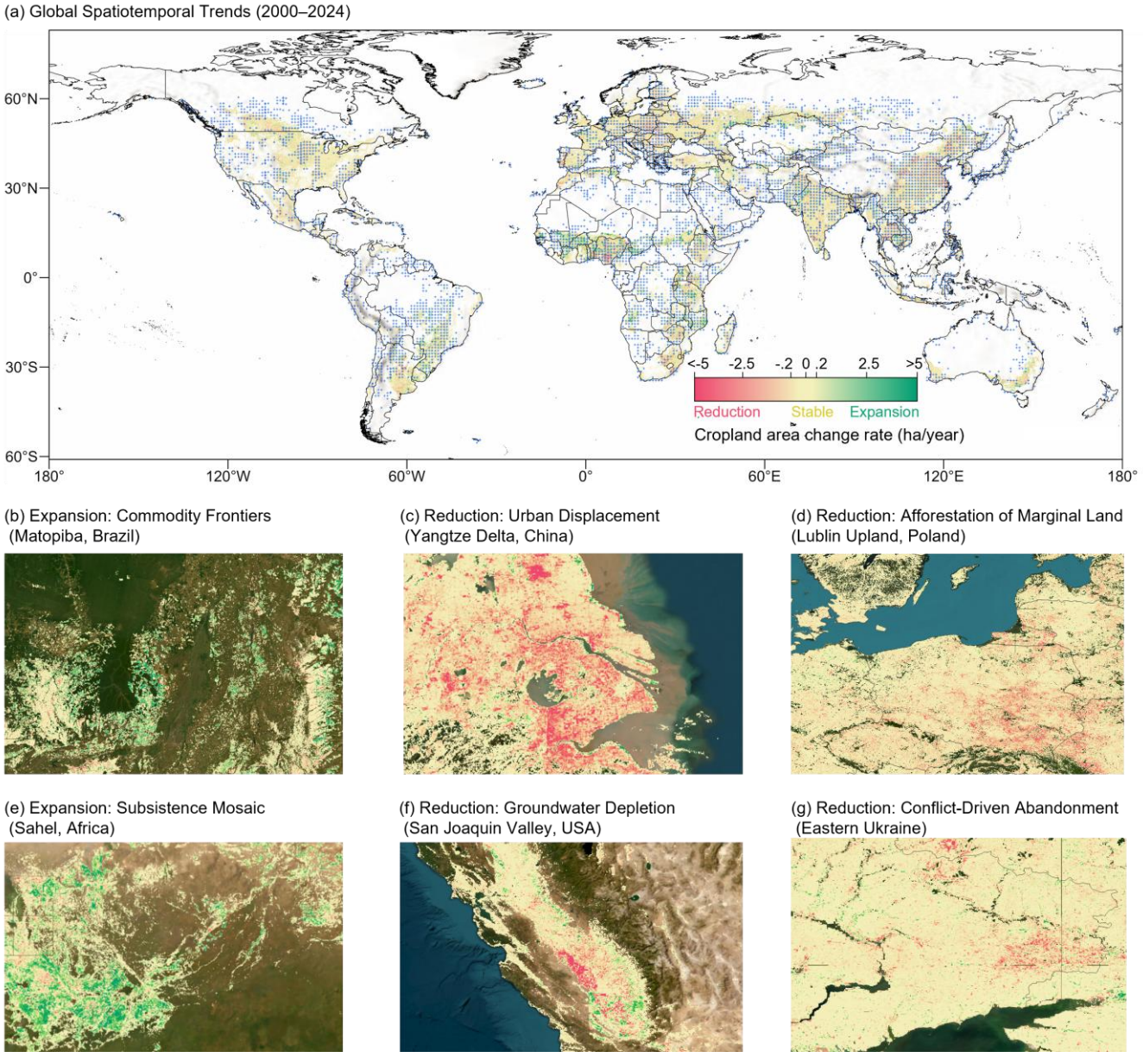


Figure 8. Spatiotemporal patterns of global cropland dynamics (2000–2024). (a) Global distribution of cropland trends aggregated at 1 km grid cells. Trends are categorized based on the Mann-Kendall test as Expansion (green, slope > 0.2 ha/year), Reduction (red, slope < -0.2 ha/year), and Stable (yellow). Blue dots indicate statistical significance ($p < 0.01$). (b–
620 h) Regional subsets illustrating diverse change drivers: (b) Commodity-driven expansion in Matopiba, Brazil; (c) Urban displacement in the Yangtze River Delta, China; (d) Afforestation-driven reduction in the Lublin Upland, Poland; (e) Subsistence expansion in the Sahel; (f) Groundwater-driven reduction in the San Joaquin Valley, USA; (g) Conflict-driven

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

abandonment in Eastern Ukraine. Background satellite imagery is sourced from Esri World Imagery Wayback (June 2020) (Sources: Esri | Powered by Esri). The national administrative boundaries used in this map are sourced from the Resources and Environmental Science Data Platform.

Leveraging the continuous annual time series of GACED30, we quantified global cropland dynamics using a rigorous statistical framework rather than simple bi-temporal differencing. As illustrated in Fig. 8a, this approach reveals a nuanced global landscape where stability is the dominant characteristic—a finding that directly corroborates the conservative area estimates discussed in Sect. 4.3.1 The vast majority of global croplands are classified as "Stable" (yellow pixels), particularly in mature agricultural regions such as the US Corn Belt, the North China Plain, and Western Europe. This spatial prevalence of stability provides the physical evidence underpinning our finding that global net cropland area has grown only marginally over the past 25 years. It further aligns with the trends observed in FAO statistics, which indicate that arable land area in developed regions has plateaued or slightly declined due to sustainable intensification, rather than expanding linearly as suggested by other satellite-derived products.

Where statistically significant change does occur, it reflects a stark "Global North-South" divergence. Expansion (green pixels) is spatially concentrated in the Global South, yet driven by divergent mechanisms. Fig. 8b highlights the "Matopiba" region in Brazil, where the Cerrado biome is rapidly being converted into large-scale commodity row crops. In contrast, the expansion observed in the African Sahel (Fig. 8e) is driven by the extensification of smallholder subsistence farming to meet population demands in semi-arid zones. Crucially, GACED30 identifies specific "hotspots" of Reduction (red pixels) that are often overlooked by optimistic models or masked by rotational fallow. Our sub-regional analysis reveals that this loss is not uniform but stems from three distinct structural drivers. In the Yangtze River Delta of East Asia (Fig. 8c), significant cropland loss is driven by urbanization, specifically the irreversible conversion of fertile peri-urban land to impervious surfaces. In the San Joaquin Valley, USA (Fig. 8f), the driver is resource constraints; unlike temporary fallow, this represents long-term land retirement driven by severe groundwater depletion and regulatory restrictions like SGMA. Meanwhile, in Eastern Europe, the reduction in Poland's Lublin Upland (Fig. 8d) and Eastern Ukraine (Fig. 8g) reflects socio-political shifts—the former due to the abandonment of fragmented strip-farms under EU afforestation incentives, and the latter due to conflict-driven infrastructure damage and landmine contamination in the Donbas region.

A critical comparison with the GLAD Cropland change map highlights the distinct advantages of the GACED30 statistical trend analysis over epoch-based differencing. Both products exhibit strong spatial consistency in identifying major expansion frontiers in South America and Africa. However, significant discrepancies exist in semi-arid and transitional zones, such as parts of Central Asia and Australia. In these regions, GLAD's epoch-based comparison (e.g., 2000-2003 vs. 2016-2019) often flags pixels as "gain" or "loss" based on temporary phenological shifts or data gaps in the early Landsat archive (Potapov et al., 2022). In contrast, the Mann-Kendall test applied to the GACED30 time series finds no statistically significant trend ($p > 0.01$) in many of these areas. This indicates that what epochal maps interpret as permanent "change" often fails the test of statistical significance when the full 25-year annual trajectory is considered. By filtering out these non-

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

significant fluctuations, GACED30 provides a more conservative and reliable map of permanent agricultural change, avoiding the overestimation of dynamics that plagues bi-temporal assessments.

5. Discussion

5.1 From static mapping to structural evolution: Advantages and uncertainties

The transition from static land cover classification to continuous structural monitoring represents a fundamental shift in assessing global agricultural dynamics. Our framework addresses the limitations of "snapshot" or "epoch-based" approaches through three distinct methodological innovations that prioritize temporal consistency. First, the integration of the SDC30 provides the observational density required to resolve the "phenological pulse" of agriculture, enabling the reconstruction of annual trajectories even in cloud-prone regions where traditional composites often fail. Second, the spectral-semantic sample alignment strategy effectively resolves the scalability bottleneck in global supervised learning by harmonizing expert-annotated samples with global spatial priors, ensuring the model learns a thematically consistent definition of "active cultivation". Finally, the shift from bi-temporal change detection to statistical structural evolution (Mann-Kendall trend analysis) serves as a rigorous filter for "ephemeral noise," distinguishing genuine land use conversion from the inter-annual variability of fallow cycles.

However, this methodological rigor necessitates a critical examination of the trade-offs between "stability" and "sensitivity." GACED30 reveals a "conservative" global trend—a net expansion of only 1.2 Mha/year—which contrasts with the rapid linear growth reported by other datasets. This divergence may stem from three distinct methodological and definitional factors.

First, it reflects the "false positive accumulation" inherent in products that rely on bi-temporal change detection. In global monitoring, short-term climatic variability—such as rainfall-induced greening in semi-arid rangelands—is often misclassified as cropland expansion in snapshot products. By employing the Mann-Kendall trend test across the full 25-year density, GACED30 effectively filters out these ephemeral fluctuations. Consequently, our "Global Stability" trend offers a more precise separation of genuine structural evolution from the transient noise of the Earth system, preventing the overestimation of agricultural encroachment.

Second, the trend may be influenced by historical observational heterogeneity. The early period of the archive relies on the fusion of sparse Landsat 5/7 coverage with MODIS data, resulting in a coarser spectral aggregation compared to the Sentinel-2 era. This historical characteristic implies that the 2000 baseline likely captures the "maximum potential range" (or upper limit) of cultivation—effectively including some mixed-pixel marginal vegetation that higher-resolution modern sensors might exclude. While this phenomenon may suggest a larger initial extent, the application of the Mann-Kendall test reduces the uncertainty of structural evolution by rigorously excluding non-monotonic noise specific to the 2000–2010 epoch. As a result, the derived Structural Evolution Indicators remain robust, focusing on confirmed anthropogenic activity rather than early-period sensor variances.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Third, the divergence is driven by semantic scope and internal transitions. Unlike products that restrict "cropland" to annual herbaceous plants, GACED30 adopts the comprehensive FAO definition, which includes permanent woody crops (e.g., orchards) and agricultural structures. Consequently, transitions that other datasets might classify as "loss" (e.g., the conversion of a wheat field to an orchard) are correctly identified by GACED30 as internal transitions within the agricultural system. By internalizing these shifts rather than flagging them as land cover turnover, GACED30 avoids the inflation of change statistics, offering a more stable and agronomically consistent baseline for global food security assessment.

5.2 Implications for global sustainability and Earth system modelling

Beyond methodological consistency, GACED30 fills a critical gap in diagnosing the structural evolution of the Anthropocene. By separating genuine land use conversion from inter-annual noise, the dataset offers three distinct analytical advantages for Earth system science.

GACED30 provides rigorous evidence for a "Global North-South Divergence," challenging narratives of uniform expansion. Unlike general products that may obscure regional trends through aggregation, our analysis confirms an anisotropic expansion of the agricultural planetary boundary. We observe a decisive shift of the expansion frontier to the Global South (driven by commodity and subsistence needs), contrasting with widespread stability or policy-driven contraction in the Global North. This spatial clarity supports the validation of "land sparing" hypotheses and helps differentiate between yield-driven stability in established regions and area-driven conversion in biodiversity hotspots.

The annual temporal fidelity of GACED30 enhances the monitoring of telecoupled systems by enabling a more precise reconstruction of land use history. Previous epoch-based products often blur transition timing, complicating the linkage between land cover change and high-frequency socioeconomic shocks. By resolving the specific year of conversion, GACED30 facilitates the causal attribution of dynamics to immediate drivers—such as price spikes, infrastructure development, or geopolitical conflict. This capability advances the field from simple change detection toward a clearer understanding of how agricultural frontiers respond to external pressures.

Finally, the "structural stability" inherent in GACED30 addresses a persistent challenge in biogeochemical modelling: the overestimation of emissions due to classification artifacts. High-resolution maps frequently suffer from "flickering," where temporary phenological variability is misclassified as rapid land turnover. By filtering rotational fallow cycles from permanent structural evolution, GACED30 offers a potentially more reliable baseline for carbon accounting. This structural consistency could ensure that modelled carbon fluxes are attributed to genuine anthropogenic encroachment rather than ephemeral noise, supporting more accurate simulations of land use–climate feedbacks.

5.3 Limitations and future work

Despite the advancements GACED30 offers in continuous global monitoring, specific limitations regarding semantic granularity remain. While including temporary fallow successfully bridges the definition gap with FAO statistics, the current dataset relies on a binary classification that aggregates active cultivation and rotational resting into a single "Cropland"

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

720 category. As noted in recent reviews, moving beyond binary classifications to characterize the specific "active status" of land is critical for accurate carbon accounting and yield modelling (Zhang et al., 2025a).

Future development will focus on three key directions. First, aligning with the vision for next-generation products, we aim to advance spatiotemporal precision by integrating 10-m Sentinel-2 imagery and in-season phenological metrics. This will allow for the resolution of smallholder plots and the characterization of multi-cropping intensities currently obscured in
725 annual composites.

Finally, we will expand the application of our Structural Evolution Indicator. Future work will move beyond simple trend detection to quantitatively extract driver-specific information—distinguishing, for example, between urbanization-driven loss and policy-driven abandonment. By parameterizing these structural changes, we aim to apply these indicators to agricultural economic modelling and policy evaluation, transforming pixel-level evolution patterns into actionable
730 socioeconomic insights for global sustainability.

6. Conclusion

Accurate quantification of the structural evolution of global agricultural systems is critical for assessing food security and monitoring planetary boundaries. Addressing the limitations of existing snapshot-based or temporally aggregated products, this study generated the first Global 30-m Annual Cropland Extent Dynamics (GACED30) dataset spanning 2000 to 2024.
735 By integrating the gap-free SDC30 with a spectral-semantic sample alignment strategy, our continuous mapping framework effectively captures the intra-annual phenological transitions of active cultivation. Comprehensive validation demonstrates that GACED30 achieves a high overall accuracy of 96.5% and an F1 score of 0.844, significantly outperforming state-of-the-art global products such as GLAD Cropland and ESA WorldCover in terms of temporal stability and thematic fidelity. Crucially, the dataset exhibits strong agreement with FAO national statistics (R^2=0.95), successfully bridging the gap
740 between remote sensing observations and official census data. Based on this consistent baseline, we estimated the global cropland area at 1488.5 Mha in 2024, revealing a distinct divergence in structural evolution: expansion is heavily concentrated in the commodity frontiers of the Global South (e.g., Africa and South America), whereas the Global North is characterized by widespread stability or policy-driven contraction. GACED30 thus provides a reliable, high-resolution evidence base for monitoring the changing footprint of global agriculture and diagnosing the long-term health of the global
745 land system.

Data availability

All datasets produced and used in this study are freely accessible from their respective repositories.

- GACED30 is publicly available at https://doi.org/10.5281/zenodo.18199675 (Chen et al., 2026).
- SDC30, and FAST sample set are available at https://data-starcloud.pcl.ac.cn/.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

- NASADEM: The digital elevation model can be accessed via the NASA Earthdata portal at https://www.earthdata.nasa.gov/data/catalog/lpcloud-nasadem-hgt-001.
- GLADCrop: The GLAD Cropland product and its associated validation sample set are available for download at https://glad.umd.edu/dataset/croplands .
- WorldCereal: The WorldCereal 2021 products can be accessed from the Zenodo repository at https://zenodo.org/records/7875105
- WorldCereal validation sample set: The validation data for WorldCereal is available at https://zenodo.org/records/7584463
- GCEP30: The GCEP30 product and associated publication can be found at the USGS publications portal https://pubs.usgs.gov/publication/pp1868
- FAOSTAT: The national agricultural statistics are publicly available from the FAO at https://www.fao.org/statistics/en

Author contribution

Yuanhong Liao: Investigation, Validation, Visualization, Writing – original draft. **Shuang Chen:** Data curation, Investigation, Methodology, Writing – original draft. **Yuqi Bai:** Conceptualization, Funding acquisition, Resources, Validation, Supervision, Writing – review & editing. **Jie Wang:** Conceptualization, Resources, Validation, Supervision, Writing – review & editing. **Peng Gong:** Conceptualization, Resources, Supervision, Writing – review & editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by National Natural Science Foundation of China (42090015).

References

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., PD Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., and R. Verón, S.: Remote sensing and cropping practices: A review, Remote Sensing, 10, 99, 2018.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., and Guanter, L.: A unified vegetation index for quantifying the terrestrial biosphere, Science Advances, 7, eabc7447, 2021.

Ceauşu, S., Leclère, D., and Newbold, T.: Geography and availability of natural habitat determine whether cropland intensification or expansion is more detrimental to biodiversity, Nature Ecology & Evolution, 1-16, 2025.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Chen, S., Wang, J., and Gong, P.: ROBOT: A spatiotemporal fusion model toward seamless data cube for global remote sensing applications, Remote Sensing of Environment, 294, 113616, 2023.
Chen, S., Liao, Y., Bai, Y., Wang, J., and Gong, P.: Global 30-m Annual Cropland Extent Dynamics (GACED30) (2000-2024) (1.0) [dataset], 10.5281/zenodo.18199675, 2026.
Chen, S., Wang, J., Liu, Q., Liang, X., Liu, R., Qin, P., Yuan, J., Wei, J., Yuan, S., and Huang, H.: Global 30-m seamless data cube (2000–2022) of land surface reflectance generated from Landsat-5, 7, 8, 9 and MODIS Terra constellations, Earth System Science Data Discussions, 2024, 1-44, 2024.
Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., and Justice, C.: The Harmonized Landsat and Sentinel-2 surface reflectance data set, Remote sensing of environment, 219, 145-161, 2018.
Damianidis, C., Santiago-Freijanes, J. J., den Herder, M., Burgess, P., Mosquera-Losada, M. R., Graves, A., Papadopoulos, A., Pisanelli, A., Camilli, F., and Rois-Díaz, M.: Agroforestry as a sustainable land use option to reduce wildfires risk in European Mediterranean areas, Agroforestry Systems, 95, 919-929, 2021.
Di Gregorio, A.: Land cover classification system: classification concepts and user manual: LCCS, Food & Agriculture Org.2005.
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., and Martimort, P.: Sentinel-2: ESA's optical high-resolution mission for GMES operational services, Remote sensing of Environment, 120, 25-36, 2012.
FAO: The future of food and agriculture – Trends and challenges, Rome, 2017.
FAO: Land use statistics and indicators. Global, regional and country trends, 2000–2020, Rome, 2022.
FAO: Tracking progress on food and agriculture-related SDG indicators 2025, Rome, 2025.
Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., and West, P. C.: Solutions for a cultivated planet, Nature, 478, 337-342, 2011.
Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., and Toulmin, C.: Food security: the challenge of feeding 9 billion people, science, 327, 812-818, 2010.
Gong, P., Wang, J., and Huang, H.: Stable classification with limited samples in global land cover mapping: Theory and experiments, Sci. Bull, 69, 1862-1865, 2024.
Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., and Bai, Y.: Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017, Sci. Bull, 64, 370-373, 2019.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote sensing of Environment, 202, 18-27, 2017.
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, 2021 IEEE international geoscience and remote sensing symposium IGARSS, 4704-4707,
Killough, B.: Overview of the open data cube initiative, IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium, 8629-8632,
Killough, B., Siqueira, A., and Dyke, G.: Advancements in the open data cube and analysis ready data—past, present and future, IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, 3373-3375,
Li, C., Gong, P., Wang, J., Zhu, Z., Biging, G. S., Yuan, C., Hu, T., Zhang, H., Wang, Q., and Li, X.: The first all-season sample set for mapping global land cover with Landsat-8 data, Science Bulletin, 62, 508-515, 2017.
Liang, X., Liu, Q., Wang, J., Chen, S., and Gong, P.: Global 500 m seamless dataset (2000–2022) of land surface reflectance generated from MODIS products, Earth System Science Data Discussions, 2023, 1-40, 2023.
Liu, H., Gong, P., Wang, J., Wang, X., Ning, G., and Xu, B.: Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020-iMap World 1.0, Remote Sensing of Environment, 258, 112364, 2021.
Mueller-Warrant, G. W., Whittaker, G. W., Griffith, S. M., Banowetz, G. M., Dugger, B. D., Garcia, T. S., Giannico, G., Boyer, K. L., and McComb, B. C.: Remote sensing classification of grass seed cropping practices in western Oregon, International Journal of Remote Sensing, 32, 2451-2480, 2011.
NASA JPL: NASADEM Merged DEM Global 1 arc second V001 [dataset], 10.5067/MEASURES/NASADEM/NASADEM_HGT.001, 2020.
Northcutt, C., Jiang, L., and Chuang, I.: Confident learning: Estimating uncertainty in dataset labels, Journal of Artificial Intelligence Research, 70, 1373-1411, 2021.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

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 sensing of Environment, 148, 42-57, 2014.
Pax-Lenney, M. and Woodcock, C. E.: The effect of spatial resolution on the ability to monitor the status of agricultural lands, Remote Sensing of Environment, 61, 210-220, 1997.
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, Nature Food, 3, 19-28, 2022.
Pretty, J.: Intensification for redesigned and sustainable agricultural systems, Science, 362, eaav0294, 2018.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A.: CatBoost: unbiased boosting with categorical features, Advances in neural information processing systems, 31, 2018.
Sassen, K., Wang, Z., and Liu, D.: Global distribution of cirrus clouds from CloudSat/Cloud-Aerosol lidar and infrared pathfinder satellite observations (CALIPSO) measurements, Journal of Geophysical Research: Atmospheres, 113, 2008.
Song, X.-P., Hansen, M. C., Potapov, P., Adusei, B., Pickering, J., Adami, M., Lima, A., Zalles, V., Stehman, S. V., and Di Bella, C. M.: Massive soybean expansion in South America since 2000 and implications for conservation, Nature sustainability, 4, 784-792, 2021.
Thenkabail, P. S., Teluguntla, P. G., Xiong, J., Oliphant, A., Congalton, R. G., Ozdogan, M., Gumma, M. K., Tilton, J. C., Giri, C., and Milesi, C.: 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, US Geological Survey2330-7102, 2021.
Tilman, D., Balzer, C., Hill, J., and Befort, B. L.: Global food demand and the sustainable intensification of agriculture, Proceedings of the national academy of sciences, 108, 20260-20264, 2011.
Tu, Y., Wu, S., Chen, B., Weng, Q., Bai, Y., Yang, J., Yu, L., and Xu, B.: A 30 m annual cropland dataset of China from 1986 to 2021, Earth System Science Data, 16, 2297-2316, 2024.
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 System Science Data, 15, 4997-5015, 2023a.
Tubiello, F. N., Wanner, N., Asprooth, L., Mueller, M., Ignaciuk, A., Khan, A. A., and RoseroMoncayo, J.: Measuring progress towards sustainable agriculture, Rome, 10.4060/cb4549en, 2021.
Tubiello, F. N., Conchedda, G., Casse, L., Pengyu, H., Zhongxin, C., De Santis, G., Fritz, S., and Muchoney, D.: Measuring the world's cropland area, Nature Food, 4, 30-32, 2023b.
Tyukavina, A., Stehman, S. V., Pickens, A. H., Potapov, P., and Hansen, M. C.: Practical global sampling methods for estimating area and map accuracy of land cover and change, Remote Sensing of Environment, 324, 114714, 2025.
Van Tricht, K., Degerickx, J., Gilliams, S., Zanaga, D., Battude, M., Grosu, A., Brombacher, J., Lesiv, M., Bayas, J. C. L., and Karanam, S.: WorldCereal: a dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping, Earth System Science Data, 15, 5491-5515, 2023.
Whitcraft, A. K., Vermote, E. F., Becker-Reshef, I., and Justice, C. O.: Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations, Remote sensing of Environment, 156, 438-447, 2015.
Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E.: Opening the archive: How free data has enabled the science and monitoring promise of Landsat, Remote Sensing of Environment, 122, 2-10, 2012.
Xie, Z., Zhao, Y., Jiang, R., Zhang, M., Hammer, G., Chapman, S., Brider, J., and Potgieter, A. B.: Seasonal dynamics of fallow and cropping lands in the broadacre cropping region of Australia, Remote Sensing of Environment, 305, 114070, 2024.
Yang, Y., Bhatta, S., and Delgado Bonal, A.: Decadal Observations of Global Daytime Cloud Properties from DSCOVR-EPIC, Frontiers in Remote Sensing, 6, 1632157, 2025.
Yu, L., Wang, J., Clinton, N., Xin, Q., Zhong, L., Chen, Y., and Gong, P.: FROM-GC: 30 m global cropland extent derived through multisource data integration, International Journal of Digital Earth, 6, 521-533, 2013.
Yu, Z., Jin, X., Miao, L., and Yang, X.: A historical reconstruction of cropland in China from 1900 to 2016, Earth System Science Data, 13, 3203-3218, 2021.
Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., and Fritz, S.: ESA WorldCover 10 m 2021 v200, 2022.

https://doi.org/10.5194/essd-2025-838
Preprint. Discussion started: 27 February 2026

Zhang, C., Kerner, H., Wang, S., Hao, P., Li, Z., Hunt, K. A., Abernethy, J., Zhao, H., Gao, F., and Di, L.: Remote sensing for crop mapping: A perspective on current and future crop-specific land cover data products, Remote Sensing of Environment, 330, 114995, 2025a.

Zhang, X., Wan, W., and Estoque, R. C.: Impacts of urban and cropland expansions on natural habitats in Southeast Asia, Nature Communications, 16, 8479, 2025b.

Zhang, X., Liu, L., Zhao, T., Zhang, W., Guan, L., Bai, M., and Chen, X.: GLC_FCS10: a global 10 m land-cover dataset with a fine classification system from Sentinel-1 and Sentinel-2 time-series data in Google Earth Engine, Earth System Science Data, 17, 4039-4062, 2025c.

Zhang, X., Zhao, T., Xu, H., Liu, W., Wang, J., Chen, X., and Liu, L.: GLC_FCS30D: The first global 30-m land-cover dynamic monitoring product with a fine classification system from 1985 to 2022 using dense time-series Landsat imagery and continuous change-detection method, Earth System Science Data Discussions, 2023, 1-32, 2023.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F.: Deep learning in remote sensing: A comprehensive review and list of resources, IEEE geoscience and remote sensing magazine, 5, 8-36, 2017.