We sincerely thank the Editors, the ESSD editorial team, and reviewers for their careful evaluation of our manuscript and their constructive comments, which have helped us substantially improve the scientific rigor, transparency, and presentation of this work. In response, we have expanded the methodological description of sample construction, CatBoost modelling, CleanLab screening, and post-classification refinement; added classifier benchmarking, threshold-sensitivity analysis, regional and agro-ecological assessments, and multi-temporal evaluations of cropland gain and loss; applied a consistent sample-based area-adjustment framework to the compared products; and clarified the effects of temporal mismatch and differing cropland definitions on product comparisons. We have also moderated overly strong claims, more explicitly acknowledged the dependence of the augmented samples on the GLAD Cropland prior and the limitations of interannual consistency and cropland-subclass discrimination, strengthened the discussion of applications and uncertainties, improved the figures and references, and carefully revised the English throughout the manuscript and Supplement. Our detailed point-by-point responses and the corresponding revisions are provided below.

Comments by Reviewer #1

General comments

This study presents a novel global 30 m resolution annual cropland dynamics dataset (GACED30) spanning 2000–2024, aiming to address the limitations of existing global cropland products in terms of temporal continuity, definitional consistency, and the monitoring of structural evolution. By constructing a gap-free SDC30 time series, introducing a spectral-semantic sample alignment strategy, and employing a CatBoost classifier combined with spatiotemporal post-processing, the authors generated the first annual, high-resolution cropland dynamics product with long-term continuity. The dataset demonstrates excellent performance in accuracy assessment and shows strong agreement with FAO national statistics, successfully revealing a “Global North-South divergence” pattern in cropland expansion and contraction. Overall, the methodology is innovative, the data product holds significant scientific value and application potential. Suggested to be published after revisions. The following are the questions and some mistakes in this manuscript.

> We sincerely thank your for the careful assessment of GACED30 and for the constructive recommendation. In response to these comments, we have

revised both the manuscript and Supplement to improve methodological transparency and reproducibility, clarify data availability and the temporal coverage of the compared products, report the sample-screening, CatBoost, and CleanLab settings, provide an explicit operational interpretation and sensitivity analysis for the minimum slope threshold, address residual temporal mismatch in inter-product comparisons, strengthen the result-based discussion, improve the cartographic elements, standardize dataset and software references, and revise the English expression throughout. We have also moderated several overly strong statements and now distinguish more carefully among static thematic agreement, transition-detection performance, national-scale statistical consistency, and the remaining limitations of the dataset. These revisions have made the presentation more precise, reproducible, and scientifically bounded.

Comments

(1) Consider adding a brief statement on data availability at the end of the abstract for completeness.

> Thank you for this helpful suggestion. Although a brief repository link was already present in the final sentence of the Abstract, we agree that the identity of the data product and the location of its archived record should be stated as explicitly as possible. We therefore revised this sentence to name GACED30 directly and provide its persistent Zenodo DOI.

The manuscript has been revised as follows:

A. Abstract, Lines 13–29 – We revised the final sentence to identify GACED30 explicitly and provide its persistent repository link.

"GACED30 provides a harmonized baseline for monitoring global cropland-extent dynamics and is publicly available at https://doi.org/10.5281/zenodo.18199675 (Chen et al., 2026)."

(2) Table 1 is comprehensive, but consider adding a column indicating the temporal coverage of each product to facilitate direct comparison with

GACED30's 25-year record.

> Thank you for this helpful suggestion. Temporal coverage was already included in the original “Temporal attributes” column. To make this information more explicit and facilitate direct comparison across products, we renamed the column as “Temporal frequency / coverage” and standardized its entries to report both the update frequency and coverage period.”

(3) The preliminary CatBoost model used for sample screening lacks detailed parameter specifications. Providing these details or including them in supplementary materials would enhance reproducibility. The CleanLab-based temporal cleaning process is mentioned but not quantitatively described. What confidence thresholds were used? How were temporal inconsistencies identified? These should be clarified.

> Thank you for this important comment. We agree that the original manuscript did not provide sufficient implementation details for either the preliminary CatBoost-based agreement screening or the subsequent CleanLab screening. We also agree that the original expression “temporal cleaning” could incorrectly imply that CleanLab directly evaluates transitions or imposes a temporal-consistency rule between adjacent years. We have therefore reorganized the description into a four-stage sample-construction procedure, reported the quantitative settings in Sect. S1.1 and Table S1, and replaced the ambiguous terminology with “year-specific weak-label screening.”

The preliminary agreement-screening classifier, the CatBoost models used to generate out-of-fold probabilities for CleanLab, and the final annual classifiers shared the following core configuration: 2,000 iterations, a tree depth of 8, a learning rate of 0.05, an L2 leaf-regularization coefficient of 6.0, Logloss as the loss function, F1 as the evaluation metric, a random seed of 42, 80 CPU threads, and CPU execution. The role of the preliminary classifier is now identified explicitly as Stage 1: it was trained exclusively with the expert-annotated FAST-Crop samples and retained a candidate only when its predicted Cropland/Non-cropland class agreed with the corresponding GLAD-

stratified label.

We further added a detailed quantitative description of Stage 2, which performs spatial, spectral, and category filtering after the preliminary agreement screening. The land surface was divided into 5°×5° geographic grids. Within each grid, the six SDC30 surface-reflectance bands were transformed as $z = \log(1 + SR)$, and FAST-Crop Cropland samples were used to estimate the corresponding multivariate Gaussian reference distribution. An augmented Cropland candidate passed the hard spectral prefilter only when its squared Mahalanobis distance satisfied $D^2 \le \chi^2(0.999, 6) = 22.457744$. Candidates were rejected when no reliable exact-grid, parent-grid, or global reference distribution was available, when D^2 exceeded this threshold, or when the log-likelihood was non-finite. Passing candidates were sampled according to their Gaussian likelihood using a likelihood temperature of 1, after which category balancing was conducted within each grid according to the Cropland/Non-cropland proportions represented by FAST-Crop.

For Stage 4, each retained augmented location was assigned annual weak labels for 2000–2024 using the relevant GLAD Cropland epoch. For each year, five-fold out-of-fold CatBoost probabilities were generated using StratifiedKFold(n_splits=5, shuffle=False) and supplied to CleanLab's default find_label_issues procedure. FAST-Crop labels were treated as fixed reference labels and were excluded from pruning, whereas only augmented point-year labels were eligible for screening. No manually selected probability or confidence threshold was applied. A flagged augmented point-year was excluded from the training sample set of the corresponding annual classifier, and an augmented location was removed entirely when more than 25% of its 25 annual labels were flagged, corresponding to at least seven years. After this procedure, the final augmented sample set contained 332,471 locations, including approximately 25,000 Cropland and 307,000 Non-cropland samples. CleanLab therefore identifies year-specific weak-label conflicts, in which the assigned GLAD-derived label is not supported by the spectral evidence for that

year; it does not identify a temporal inconsistency through an explicit year-to-year transition rule. Direct temporal information is introduced separately during post-classification refinement through the limited rotation rule in Sect. 3.3.1. We have revised the manuscript accordingly to distinguish sample screening from temporal refinement.

The manuscript has been revised as follows:

A. Sect. 3.2, Lines 269–273 – We replaced the previous “temporal refinement” description with year-specific weak-label screening and clarified that direct temporal linkage is introduced only during post-classification refinement.

“To generate annual cropland probability maps under a common mapping framework, we first constructed spatially extensive and thematically harmonized annual training subsets through multi-stage sample alignment and year-specific weak-label screening (Sect. 3.2.1). We subsequently used year-specific CatBoost models to estimate pixel-wise cropland probabilities across the 2000–2024 archive (Sect. 3.2.2). The annual classifiers were trained independently, and direct temporal linkage between their outputs was introduced only through the limited post-classification rotation rule described in Sect. 3.3.1.”

B. Sect. 3.2.1, Lines 285–294 – We added the main-text description of Stage 2 spatial, spectral, and category filtering and Stage 3 performance-guided sample optimization, with detailed settings directed to Sect. S1.1 and Table S1.

“In Stage 2, spatial, spectral, and category filtering was performed. The global land surface was divided into 5° × 5° geographical grids. Within each grid, augmented candidates were sampled according to the statistical distribution of the six-band annual-median SDC30 surface-reflectance vector, drawing on the approximately Gaussian behaviour of log-transformed surface reflectance in homogeneous land-cover regions (Liao et al., 2026a). Category balancing was then performed by sampling the filtered augmented candidates according to the Cropland/Non-cropland proportions of FAST-Crop within each grid. In Stage 3,

the remaining candidates were optimized using the sample-size–performance scaling relationship established by Liao et al. (2026b). At each iteration, alternative 10,000-sample batches were evaluated, and the batch producing the highest estimated asymptotic performance was retained. Iterations were stopped once the retained augmented-sample total reached the target of approximately 340,000. Detailed implementation settings are provided in Sect. S1.1 and Table S1."

C. Sect. 3.2.1, Lines 301–314 – We added the annual weak-label assignment, five-fold out-of-fold prediction design, CleanLab operation, pruning scope, point-year exclusion rule, and location-level removal criterion.

"Finally, annual weak labels were assigned directly to the selected spatial samples for 2000–2024. For 2000–2019, the label from each GLAD Cropland epoch was assigned to every year covered by that epoch; because GLAD Cropland provides no subsequent epoch, the 2016–2019 labels were also used as the initial weak labels for 2020–2024. For each year, five-fold out-of-fold CatBoost probabilities were generated using StratifiedKFold(n_splits=5, shuffle=False) and supplied to CleanLab's default find_label_issues procedure (Northcutt et al., 2021). FAST-Crop labels were treated as fixed reference labels and were excluded from possible pruning, whereas only the augmented point-year labels were eligible for screening. No manually selected probability or confidence threshold was applied. An augmented point-year flagged by CleanLab was excluded from the training sample set for the corresponding annual classifier, and an augmented location was removed entirely when more than 25% of its 25 annual labels were flagged, corresponding to at least seven years. This procedure screens year-specific weak-label conflicts rather than imposing a transition rule between adjacent years. The annual training subsets therefore shared the same expert-reference basis, feature definition, sample-construction framework, and screening procedure, but were not identical because augmented point-year labels could be excluded separately for each

annual classifier. After annual weak-label assignment and screening, the final augmented sample set contained 332,471 locations, comprising approximately 25,000 Cropland and 307,000 Non-cropland samples."

D. Sect. 3.2.2, Lines 316–320 – We clarified that the preliminary screening classifier, the out-of-fold CatBoost models used for CleanLab, and the final annual classifiers shared the same reported core configuration.

"Following construction of the annual training subsets, we trained independent year-specific classifiers to account for interannual phenological variability. For each year from 2000 to 2024, the corresponding valid FAST-Crop and augmented samples were used to train a CatBoost classifier (Prokhorenkova et al., 2018). The preliminary screening classifier, the models used to generate out-of-fold probabilities for CleanLab, and the final annual classifiers shared the same core CatBoost configuration, which is reported in Sect. S1.1 and Table S1."

E. Sect. 3.3.1, Lines 344–352 – We clarified that temporal information is incorporated after annual classification through a limited rotation rule rather than through CleanLab.

"Because the annual CatBoost models were trained independently, temporal evidence was introduced after classification through a limited rotation rule targeting uncertain observations. Specifically, we identified pixels within the probability interval ($P \in [0.3, 0.5]$), which would otherwise be classified as Non-cropland under the standard binary threshold.

For a pixel in year t, an uncertain observation within this interval was retained as Cropland under a temporary-fallow interpretation only when the same pixel was classified as active cropland ($P>0.5$) in both the preceding year t-1 and succeeding year t+1. This rule bridges isolated one-year gaps that are compatible with temporary fallow or weak phenological signals. The temporal context reduces, but does not eliminate, confusion with spectrally similar

pasture, rangeland, or bare surfaces. The rule does not alter observations with probabilities below 0.3, smooth complete pixel trajectories, or impose general temporal consistency across the annual series."

F. Sect. S1.1, Lines 15–59 – We added the detailed four-stage construction procedure. In particular, Stage 1 specifies the preliminary CatBoost agreement screening, Stage 2 reports the 5°×5° spatial stratification, six-band transformation, Gaussian reference distribution, Mahalanobis-distance threshold, likelihood sampling, and category-balancing rules, and Stage 4 specifies the annual CleanLab screening.

G. Table S1, Lines 60 – We added the complete CatBoost configuration together with the principal Stage 2 filtering, sample-optimization, and CleanLab settings.

(4) The selection of β = 0.2 ha/year as the threshold for significant change (within a 1 km grid) lacks explicit justification. Please provide reasoning for this threshold (e.g., based on expected noise levels or agronomic considerations).

> Thank you for identifying that the original justification of the threshold was insufficient. We agree that the previous statement—that 0.2 ha yr^{-1} was selected to filter out fluctuations caused by temporary fallowing or inter-annual classification instability—was not adequately supported by an explicit noise calibration or an agronomic benchmark. We have therefore removed this unsupported claim and substantially revised Sect. 3.4.

First, we now distinguish the estimated Theil–Sen slope β from the selected minimum effect-size threshold τ. Statistical significance is determined by the Mann–Kendall p-value, whereas τ specifies the minimum trend magnitude interpreted as grid-scale structural change. We also distinguish the general structural trend, defined by the sign and magnitude of β, from the stringent structural trend, which additionally requires $p<0.01$.

We further clarify that neither the Mann–Kendall test nor the Theil–Sen

estimator provides a theoretically unique or universally optimal value of τ. In this study, τ=0.2 ha yr^{-1} was adopted as an operational threshold with an interpretable grid-scale meaning. A 1×1 km grid covers approximately 100 ha, so this value represents a linear trend equivalent to 0.2% of the grid area per year and 4.8 ha, or 4.8% of the grid area, across the 24 annual intervals between 2000 and 2024. Although 0.2 ha is numerically equivalent to approximately 2.2 nominal 30 m pixel areas, τ is estimated from the complete 25-year trajectory rather than applied as a single-year pixel-count threshold.

To make the dependence on this operational choice transparent, we added sensitivity analyses for τ=0.10–0.50 ha yr^{-1} at intervals of 0.05 ha yr^{-1}. Increasing τ progressively reassigns lower-magnitude expansion and reduction grid cells to stable, as expected. Nevertheless, the continental composition of the detected changes varies only modestly, and nearby thresholds retain high spatial overlap with the τ=0.2 reference masks. At τ=0.15 and 0.25, the Jaccard indices are approximately 0.88–0.91 for the general masks and 0.91–0.93 for the stringent masks. These results support τ=0.2 as a reasonable and interpretable operational reference for this analysis, without implying that it is uniquely optimal or universally applicable.

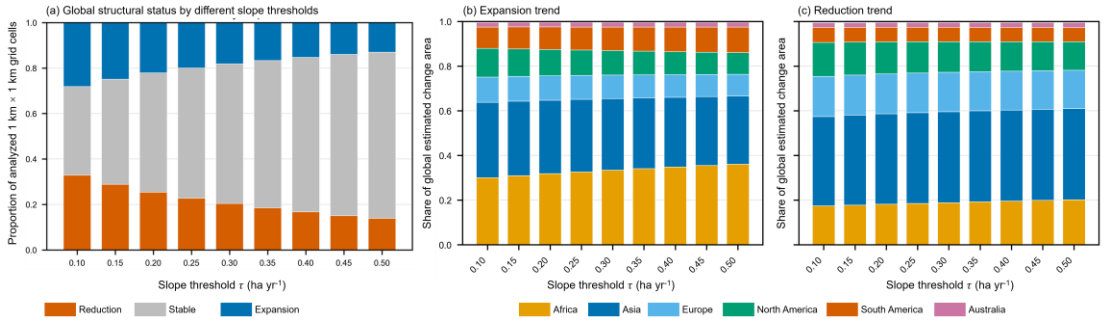

Figure S1. Sensitivity of global structural evolution trend composition and the continental distribution of cropland change to the minimum slope threshold. (a) Proportions of analysed 1 × 1 km grid cells assigned to general expansion (β>τ), general reduction (β<−τ), and general stable (|β|≤τ) as τ varies from 0.10 to 0.50 ha yr−1. This panel applies the general structural trend definition independently of MK significance. (b–c) Continental shares of the total global area classified as expansion and reduction, respectively, under the corresponding thresholds. For each threshold, the continental shares within each change class sum to 1. Panels (b–c) describe the continental composition of the globally detected

change area.

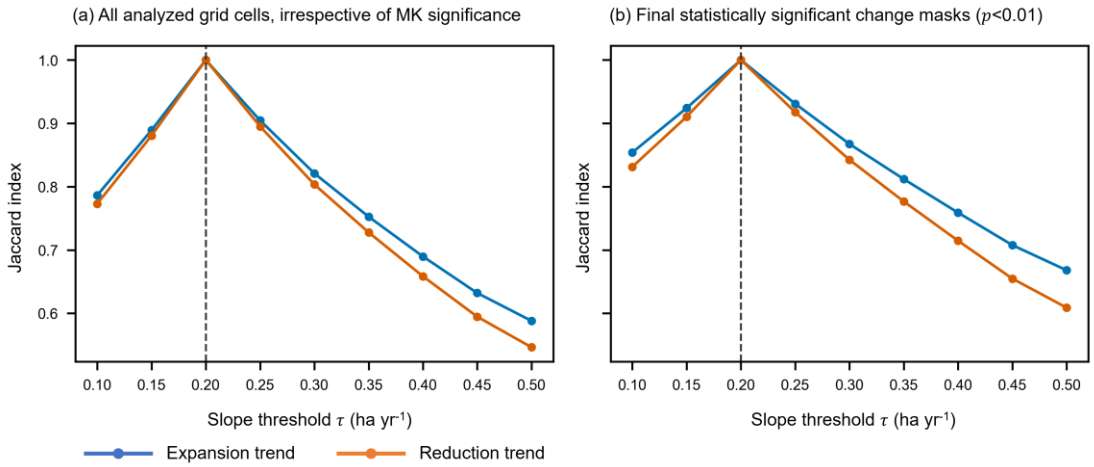

Figure S2. Sensitivity of the spatial expansion and reduction masks to the minimum slope threshold. The conventional Jaccard index, $J=|M_\tau \cap M_{0.2}| / |M_\tau \cup M_{0.2}|$, compares the mask obtained at each τ with the corresponding reference mask at τ=0.2. Results are shown for (a) all analysed 1 × 1 km grid cells classified according to the sign and magnitude of the Theil–Sen slope, irrespective of Mann–Kendall significance, and (b) the final statistically significant change masks, for which expansion requires p<0.01 and β>τ, and reduction requires p<0.01 and β<−τ. A Jaccard index of 1 indicates identical masks.

The manuscript has been revised as follows:

A. Sect. 3.4, Lines 373–394 – We replaced the original threshold passage and revised the structural-trend definition.

"To resolve this ambiguity and identify the directional evolution of these landscapes, we derived Structural Evolution Indicators from GACED30 using a grid-based trend analysis framework (Fig. 2d). Area calculations were performed in a global equal-area projection following the practical principles outlined by Tyukavina et al. (2025). The annual 30 m binary cropland maps were aggregated to 1 × 1 km grid cells, and mapped cropland area was summed within each grid cell for each year. This produced a 25-observation annual cropland-area series for each analysed grid cell over 2000–2024. The non-parametric Mann–Kendall (MK) test and Theil–Sen estimator were applied to each time series, yielding the MK significance level (p), the Theil–Sen slope (β, ha yr^{-1}), and the confidence interval of the slope. Using the minimum slope threshold τ, grid cells with β>τ, β<−τ, and $|\beta| \leqslant \tau$ were classified as general

expansion, general reduction, and stable, respectively. Expansion and reduction were additionally classified as stringent when the corresponding Mann–Kendall test satisfied p<0.01.

Neither the MK test nor the Theil–Sen estimator provides a theoretically unique or universally optimal value of τ. The appropriate minimum effect size depends on the spatial aggregation unit, study duration, and analytical objective. For the principal analysis in this study, we adopted τ=0.2 ha yr^{-1} as an operational grid-scale threshold. Because a 1 × 1 km grid cell covers approximately 100 ha, this value corresponds to a linear trend equivalent to 0.2% of the grid area per year and to 4.8 ha, or 4.8% of the grid area, across the 24 elapsed annual intervals between 2000 and 2024. It therefore provides an interpretable minimum magnitude for identifying landscape-scale structural change over the study period.

Although 0.2 ha is numerically equivalent to approximately 2.2 nominal 30 m pixel areas, τ is not applied as a single-year pixel-count detection limit. Instead, β is estimated from the complete 25-year cropland-area trajectory within each grid cell. We do not consider τ=0.2 ha yr^{-1} to be universally optimal, and other thresholds may be appropriate for analyses with different objectives or tolerances for lower-magnitude change. We therefore evaluated the sensitivity of both the general and stringent structural-evolution classifications by varying (τ) from 0.10 to 0.50 ha yr^{-1} at intervals of 0.05 ha yr-1 (Figs. S1 and S2).”

B. Sect. 4.3.3, Lines 767–774 – We revised the opening results paragraph to report both the general and stringent structural-trend definitions and the class composition under τ=0.2 ha yr^{-1}.

“Leveraging the continuous annual time series of GACED30, we quantified global cropland dynamics through grid-level MK and Theil–Sen trend analysis rather than simple bi-temporal differencing. Fig. 8a presents both complementary interpretations of the resulting trends: the general structural

trend describes all grid cells whose estimated slope exceeds the selected minimum magnitude, whereas the stringent structural trend identifies the subset that additionally satisfies p<0.01. Under the principal threshold of τ=0.2 ha yr^{-1}, stable was the dominant structural-status class, accounting for slightly more than half of the analyzed grid cells; approximately one-quarter were classified as general reduction and slightly more than one-fifth as general expansion (Fig. S1a). Stable cropland was especially extensive in established agricultural regions such as the US Corn Belt, the North China Plain, and Western Europe."

C. Sect. 4.3.3, Lines 775–786 – We added a separate paragraph reporting the threshold-sensitivity results, including the progressive reassignment of change cells to stable, continental composition, and Jaccard overlap with the τ=0.2 masks.

"Because τ is an operational rather than theoretically optimal threshold, we examined whether the principal geographical interpretation depended strongly on its selected value. Increasing τ from 0.10 to 0.50 ha yr^{-1} progressively reassigned lower-magnitude expansion and reduction grid cells to stable, as expected, and therefore changed the absolute mapped extent of structural change (Fig. S1a). However, the continental composition of the general expansion and reduction areas changed only modestly across the evaluated thresholds, with no abrupt redistribution around τ=0.2 ha yr^{-1} (Fig. S1b–c). Spatial agreement with the τ=0.2 reference masks was also high for nearby thresholds under both definitions: at τ=0.15 and 0.25, the Jaccard indices were approximately 0.88–0.91 for the general masks and 0.91–0.93 for the stringent masks (Fig. S2). Thus, the precise extent of lower-magnitude change remains threshold-dependent, but the broad continental distribution and principal spatial patterns are retained across reasonable values surrounding τ=0.2. Together with the regional patterns in Fig. 8, this supports the robustness of the central interpretation that agricultural expansion is disproportionately concentrated in the Global South, whereas widespread stability and localized contraction

characterize much of the Global North. Nevertheless, the sensitivity results do not imply that τ=0.2 is a uniquely optimal threshold."

D. Sect. 3.4, Lines 395–399 – We revised the methodological qualification of the complete annual trajectory to avoid claiming that trend analysis guarantees a permanent conversion or independently determines its cause.

"Compared with bi-temporal differences between multi-year epochs, analysis of all 25 annual observations reduces sensitivity to the selection of baseline and endpoint windows and limits the influence of short-lived fluctuations. The Mann–Kendall test and Theil–Sen estimator quantify the statistical support, direction, and magnitude of mapped trends under the specified criteria. These statistics do not independently establish the cause of a trend or guarantee that every mapped change represents a permanent land-cover conversion."

E. Sect. 5.1, Lines 813–819 – We revised the Discussion to state that the full trajectory reduces sensitivity to endpoint selection while the exact spatial extent of expansion and reduction remains threshold-dependent.

"The complete annual record supports two complementary descriptions of cropland dynamics. Cropping Frequency summarizes the proportion of years in which a pixel was mapped as Cropland, whereas the Structural Evolution Indicators use annual grid-level cropland area, the Mann–Kendall test, and Theil–Sen slopes to identify persistent directional patterns. Analysis of the full trajectory reduces sensitivity to the selection of individual endpoint years, but the exact spatial extent of expansion and reduction remains dependent on the minimum slope threshold. The sensitivity analysis indicates that the broad continental distribution is retained across thresholds near τ=0.2 ha yr^{-1}, without implying that this value is uniquely optimal or that the detected trends establish their underlying causes."

(5) The comparison with other products (Table 2) is valuable, but some products (e.g., GLAD Cropland, ESA World Cover) have differing temporal baselines.

The potential impact of temporal mismatches on comparative metrics should be explicitly addressed.

> Thank you for highlighting this important limitation. We agree that differing temporal baselines mean that the comparison in Table 2 is not perfectly contemporaneous and that nearest-date matching can reduce, but cannot eliminate, the resulting temporal offset. We also acknowledge that the original wording could be read as overstating the relative accuracy of GACED30. We have therefore removed language implying statistically significant or universal superiority over all peer products. The revised manuscript interprets Table 2 as evidence of relative static thematic agreement with the adopted FAST-Crop/GACED30 reference definition, rather than as proof that GACED30 is universally more accurate across all cropland categories, years, or change conditions. This limitation is stated directly in the interpretation of Table 2 in Sect. 4.1.1 and is carried into the broader discussion of incomplete temporal comparability and reference coverage in Sect. 5.3.

The FAST-Crop validation samples predominantly represent land-cover conditions around 2015. For each product, we therefore selected the available annual layer or multi-year epoch closest to the reference-imagery year and constructed one confusion matrix for that product. This procedure minimizes temporal offset, but genuine cropland changes occurring between the reference date and the selected product layer or epoch can still be recorded as disagreement. Consequently, part of the difference among comparative metrics may reflect residual temporal mismatch in addition to mapping behaviour and cropland-definition differences.

The manuscript has been revised as follows:

A. Sect. 3.5.1, Lines 409–413 – We clarified the nearest-date matching strategy used to compare products with different temporal availability.

“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_sample-T_product |. This ensures that the validation reflects the product's performance closest to the ground reference date."

B. Sect. 4.1.1, Lines 552–558 – We explicitly acknowledged that temporal and definitional mismatches can contribute to disagreement and qualified the interpretation of Table 2 so that it is not presented as evidence of universal superiority.

"Part of the mismatch with peer products may reflect differences in cropland definitions rather than mapping errors alone. Specifically, GACED30 includes orchards and active fallow but excludes temporary meadows and pastures, whereas peer products differ in their treatment of these categories. Temporal mismatches between the validation samples and product layers may also contribute to the observed differences. The metrics in Table 2 should therefore be interpreted as the products' relative agreement with the adopted FAST-Crop/GACED30 definition rather than as evidence of universal superiority across all cropland categories. Accordingly, Table 2 supports the static thematic reliability of GACED30 under the adopted Cropland/Non-cropland reference definition but does not directly assess its ability to detect inter-period cropland changes."

C. Sect. 5.3, Lines 868–875 – We acknowledged the broader limitation that complete interannual comparability is not enforced and that available multi-temporal references do not validate every annual layer or exact transition year.

"Temporal comparability is supported by the common observation, feature, sample-construction, and model protocols, but complete pixel-level interannual consistency is not enforced. The independently trained annual classifiers can

remain sensitive to annual spectral conditions, particularly near the classification threshold or during abrupt land-cover changes. Differences between the earlier Landsat–MODIS observation environment and the more recent sensor era may also influence temporal comparability and are not eliminated solely by trend analysis. Moreover, the available GLAD Cropland and LUCAS references assess crop gain and crop loss across multi-year intervals rather than every annual layer. They cannot identify the exact transition year, detect all within-interval reversals, or distinguish permanent cropland reduction from temporary non-cropland states. More independent annual and transition-specific reference observations are therefore needed."

(6) Please further discuss based on the results of this study to enhance the depth of the discussion.

> Thank you for this helpful suggestion. We agree that the original Discussion focused primarily on the general implications of GACED30 and did not sufficiently connect those implications to the principal results of this study. We have therefore rewritten Sect. 5.2 around the modest increase in global cropland area, the predominance of stable cropland, and the spatial concentration of expansion and reduction. The revised discussion explains how the coexistence of relative global stability and substantial regional reorganisation can inform food-security assessment, agricultural-frontier monitoring, land-use policy evaluation, carbon accounting, and Earth system model benchmarking. We also state the complementary data required for these applications and avoid interpreting temporal correspondence as evidence of causality.

The manuscript has been revised as follows:

A. Sect. 5.2, Lines 833–847 – We connected the observed modest global net increase and concentrated regional reorganisation to food-security assessment, agricultural-frontier monitoring, and land-use policy analysis, while defining the limits of those applications.

"The results in Sects. 4.3.1–4.3.3 indicate that global cropland increased

modestly from 1458.5 Mha in 2000 to 1488.5 Mha in 2024, while stable cropland remained predominant and statistically significant expansion and reduction were spatially concentrated. This combination shows that a relatively small global net change can coexist with substantial regional reorganisation, particularly the contrast between concentrated expansion in parts of Africa and South America and widespread stability or reduction in much of the Global North. Because GACED30 maps cropland extent rather than crop type, yield, production, or food access, it cannot measure food security directly. When combined with crop-type maps, yield and production statistics, population, market-access, and nutrition data, however, it can provide the land-footprint component of food-security assessments and help distinguish cases in which production changes are accompanied by cropland expansion from those occurring without marked extent change. The significant trend layer can also identify candidate agricultural-frontier and persistent cropland-reduction zones for targeted monitoring; determining the preceding land cover and environmental implications requires independent forest, grassland, biodiversity, protected-area, infrastructure, and land-tenure information. When combined with trade-flow and socioeconomic data, these spatial trajectories can support telecoupled land-system analyses linking distant production and consumption regions; when paired with policy boundaries, implementation dates, and an appropriate comparison or counterfactual design, they can also serve as an outcome layer for land-use policy evaluation, although GACED30 alone cannot establish causal effects."

B. Sect. 5.2, Lines 848–857 – We explained how the structural-trend record can support carbon-accounting sensitivity analyses and Earth system model benchmarking when combined with appropriate complementary datasets.

"The structural trend record also has practical value for carbon accounting and land-use–climate analysis because persistent directional changes can be distinguished from short-term fluctuations under a transparent statistical

criterion. When combined with preceding land-cover classes, biomass and soil-carbon stock data, emission factors, and model assumptions, mapped cropland expansion or reduction can provide spatial activity information or stratification for carbon-accounting and sensitivity analyses; GACED30 itself does not quantify carbon stocks or emissions. At coarser scales, the 2000–2024 record can be aggregated and harmonised with model grids to compare the spatial allocation, direction, and persistence of observed cropland trends with land-use inputs or outputs used by Earth system models, including LUH2 (Hurtt et al., 2020). Such comparisons can support evaluation of land-use inputs and model behaviour, but GACED30 is not a complete land-use forcing dataset and cannot replace datasets that represent crop types, pasture, management, and transitions among non-cropland classes."

(7) The scale is missing in all the spatial maps. Some small maps lack the north pointer or latitude/longitude.

> Thank you for this helpful comment. We carefully checked the spatial map panels throughout the manuscript. Scale bars and north pointers have been added to the relevant regional maps.

(8) Ensure consistency in formatting, particularly for datasets and software references (e.g., NASA JPL NASADEM citation).

> Thank you for this helpful suggestion. We have carefully checked and standardised the formatting of all references throughout the manuscript, including datasets and software-related references. In particular, the NASA JPL NASADEM citation has been revised to follow the recommended dataset citation format, with consistent author, dataset, repository, persistent identifier, and year information.

(9) Please improve English expression in the manuscript.

> Thank you for this helpful suggestion. We have carefully reviewed and revised the entire manuscript to improve English expression, grammatical accuracy, clarity, and overall fluency. We also standardized terminology and phrasing where appropriate.

Comments by Reviewer #2

General comments

The authors present the first global 30 m annual cropland extent dynamics product for 2000–2024, with a clear methodological framework, independent FAST-based validation, comparison against several global products, and consistency checks against FAOSTAT. The dataset itself is highly valuable and likely to be useful for global land-system research. That said, several issues should be addressed to further strengthen the manuscript.

> Thank you very much for your positive assessment of GACED30 and for recognizing the potential value of its 25-year annual record for global land-system research. We also appreciate your constructive recommendations for strengthening the evaluation and interpretation of the dataset. In response, we added a matched regional assessment in China using CACD and the Third National Land Survey, an agro-ecological-zone-stratified accuracy assessment, a controlled classifier benchmark against Random Forest and XGBoost, and a targeted multi-temporal assessment of crop gain and crop loss using GLAD Cropland and LUCAS reference observations. We also expanded the discussion of practical applications, clarified the effects of differing cropland definitions on inter-product comparisons, explicitly identified landscapes and cropland subclasses for which performance is less well constrained, and clarified the raster coding convention. Throughout the revised manuscript, we moderated claims of comparative performance and more clearly distinguished static thematic accuracy, transition-detection evidence, regional consistency, and the remaining limitations of GACED30.

Major Comments

(1) The current validation relies mainly on the FAST validation set, comparisons with six global products, and consistency with FAOSTAT across 132 countries. This is useful, but still not enough to demonstrate accuracy in key agricultural regions, especially in areas with fragmented fields, complex cropping systems,

or intensive land-use transitions. I strongly recommend adding comparisons with high-quality regional products and independent regional statistics, for example the CACD product in China and official cropland statistics such as the Third National Land Survey.

> Thank you for this important and specific recommendation. We agree that the original global evaluation did not provide sufficient direct evidence from a major agricultural region characterized by heterogeneous cropping systems, fragmented fields, and intensive land-use change. We therefore added a China-focused assessment comprising three complementary components: a matched comparison of GACED30 and CACD against the same 2015 GLCVSS reference subset, a comparison of the 2019 GACED30 area estimate with the cultivated-land and garden-land areas reported by the Third National Land Survey, and a comparison of the 2021 sample-adjusted GACED30 and CACD estimates. Both products achieved an overall accuracy of 0.94 under the matched reference-sample protocol. The 2019 adjusted GACED30 estimate was 179.13 Mha, compared with the approximately comparable, but not definitionally identical, Third National Land Survey benchmark of 148.03 Mha, whereas the 2021 GACED30 estimate of 179.41 Mha fell within the uncertainty interval of the published CACD estimate of 172.52 ± 21.24 Mha. Because these sources differ in classification scope, observation protocols, spatial representation, and area-estimation procedures, we interpret them as complementary regional evidence rather than treating any source as error-free ground truth.

The manuscript has been revised as follows:

A. Sect. 2.5, Lines 207–222 – We added a new subsection introducing CACD, the matched Chinese GLCVSS reference subset, and the Third National Land Survey, and clarified the comparability and remaining definitional differences among these sources.

“For regional evaluation in China, we employed the China Annual Cropland Dataset (CACD), a 30-m annual cropland dataset covering 1986–2021 (Tu et al., 2024). CACD estimates annual cropland probabilities and applies LandTrendr trajectory segmentation together with a spatial–temporal consistency check to reduce interannual noise and refine its annual maps. These explicit temporal-refinement procedures make CACD a valuable regional benchmark for evaluating annual cropland time series.

For a controlled regional accuracy comparison, we used the China subset of the Global Land Cover Validation Sample Set (GLCVSS) adopted in the published CACD assessment. GLCVSS contains all-season land-cover interpretations based on Landsat 8 observations acquired between 2013 and 2015 and follows the FROM-GLC classification system (Li et al., 2017). Following Tu et al. (2024), we retained the GLCVSS to evaluate the 2015 GACED30 and CACD layers. Because both products were evaluated using the same GLCVSS subset, this analysis represents a matched regional product comparison rather than an additional independent reference source.

We further used the Third National Land Survey as an official benchmark for cropland area in China. The survey represents land-use conditions as of 31 December 2019 and reports 127.86 Mha of cultivated land and 20.17 Mha of garden land (State Council Third National Land Survey Leading Group Office et al., 2021). Because garden land includes orchards and other perennial cultivation, combining the two categories gives an area of 148.03 Mha that is more comparable, although not definitionally identical, to the GACED30 cropland class.”

B. Sect. 3.5.3, Lines 472–485 – We added the matched regional evaluation and definition-aware national-area comparison procedures.

“We conducted a matched regional assessment in China by evaluating the 2015 GACED30 and CACD layers against the same GLCVSS sample units. Using the same reference samples, target year, and binary Cropland/Non-cropland evaluation protocol minimized differences associated with sampling design and temporal matching. We calculated OA according to the definitions provided in Sect. 3.5.1. Because the comparison is based on identical reference units, we report the resulting metrics directly without reproducing separate confusion matrices. Nevertheless, differences between the native cropland definitions of GACED30 and CACD remain relevant to their interpretation.

For the national area assessment, we compared the China-specific GACED30 estimates produced using the pooled FAST-Crop area-adjustment estimator described in Sect. 3.5.4 with two complementary regional benchmarks. The adjusted GACED30 estimate for 2019 and its 95% confidence interval were compared with the combined cultivated-land and garden-land area reported by the Third National Land Survey. For 2021, we compared the adjusted GACED30 estimate with the published sample-adjusted CACD estimate and its reported 95% confidence interval (Tu et al., 2024). Because the official survey and remote-sensing products differ in classification scope, observation protocol, spatial representation, and area-estimation procedure, these comparisons were used to assess consistency rather than to treat any individual estimate as error-free ground truth.”

C. Sect. 3.5.4, Lines 511–515 – We extended the sample-adjusted area-estimation framework to China and distinguished the raw mapped area from the reference-sample-adjusted estimate.

“For the China-specific estimate, the same area-adjustment framework was applied using the pooled subset of FAST-Crop validation records located within China. The China-wide mapped Cropland and Non-cropland area proportions were combined with the corresponding regional reference error matrix to derive

the sample-adjusted cropland area, standard error, and 95% confidence interval following Olofsson et al. (2014). Raw mapped and sample-adjusted areas were retained separately to avoid conflating pixel-counted map area with the reference-sample-adjusted estimate."

D. Sect. 4.1.3, Lines 600–614 – We added the China-specific results, including the shared OA of GACED30 and CACD and the definition-aware comparisons with the Third National Land Survey and CACD.

"The matched regional evaluation in China showed that GACED30 and CACD both achieved an OA of 0.94 against the GLCVSS sample units for 2015. This result indicates comparable overall performance under the common reference-sample protocol. Because the assessment represents a single reference year and the two products differ in their native cropland definitions and temporal-processing strategies, it is not interpreted as evidence that either product is generally superior.

The raw mapped cropland area of GACED30 in China in 2019 was 168.07 Mha. Applying the pooled FAST-Crop China area-adjustment estimator yielded an estimated cropland area of 179.13 Mha, with a standard error of 2.57 Mha and a 95% confidence interval of 174.10–184.16 Mha. The Third National Land Survey reported 127.86 Mha of cultivated land and 20.17 Mha of garden land, producing an approximately comparable combined benchmark of 148.03 Mha. The adjusted GACED30 point estimate was therefore 31.10 Mha, or 21.0%, higher than this combined official value. This confidence interval quantifies sampling uncertainty in the remote-sensing area estimator but does not account for differences in classification scope, spatial representation, survey or mapping procedures, and area-estimation methods. This discrepancy should therefore not be interpreted as showing that either source provides an error-free estimate of cropland extent. In 2021, the adjusted GACED30 estimate was 179.41 Mha, compared with the published sample-adjusted CACD estimate of 172.52 ±

21.24 Mha. Their point estimates differed by 6.89 Mha, or approximately 4.0%, and the GACED30 estimate fell within the 95% confidence interval reported for CACD."

E. Sect. 4.1.4, Lines 648–652 – We replaced the previous China-specific interpretation with a neutral explanation based on classification scope, spatial measurement, reporting frameworks, sample adjustment, mixed-pixel effects, and mapping errors.

"For China, GACED30 estimates were also higher than the corresponding FAOSTAT values. As further demonstrated through the comparisons with CACD and the Third National Land Survey in Sect. 4.1.3, these differences may reflect classification scope, spatial measurement, reporting frameworks, sample adjustment, mixed-pixel effects, and mapping errors. They should therefore not be interpreted as evidence that either source systematically underestimates the true cropland area."

F. Abstract, Sect. 1, and Sect. 6, We added concise statements indicating that the global evaluation is complemented by a matched regional assessment in China.

(2) The manuscript uses CatBoost as the core classifier and provides a brief rationale for this choice, but there is no direct comparison with other widely used machine-learning models such as Random Forest or XGBoost/LightGBM. Without such benchmarking, it is difficult to judge whether CatBoost is truly the most suitable model for this task. At minimum, the authors should include one or two baseline model comparisons and report both accuracy and computational efficiency.

> Thank you for this helpful suggestion. We agree that the original manuscript justified CatBoost primarily through its algorithmic characteristics without

providing an empirical comparison with alternative classifiers. We therefore added a controlled benchmark against Random Forest and XGBoost using the same FAST-Crop plus augmented training framework and the same 557-dimensional spectro-temporal feature representation. CatBoost achieved a Macro-F1 of 0.8358, compared with 0.8324 for XGBoost and 0.8283 for Random Forest. On the same NVIDIA A100 GPU, CatBoost processed 3,148,493 samples per second compared with 719,361 samples per second for XGBoost. Random Forest processed 163,614 samples per second on a CPU platform; therefore, its throughput is reported only as an implementation-specific reference and not as a hardware-normalized comparison. We have accordingly limited our conclusion to the evaluated implementations: CatBoost provided the highest observed predictive performance and the fastest same-platform GPU inference, rather than being universally superior under all datasets or computing environments.

The manuscript has been revised as follows:

A. Sect. 3.2.2, Lines 321–325 – We expanded the rationale for selecting CatBoost and added the classifier-benchmark design.

“CatBoost was selected because its ordered-boosting mechanism mitigates prediction shift in conventional gradient-boosted decision trees, while its oblivious-tree structure efficiently handles the 557-dimensional spectro-temporal feature space without manual feature reduction. To assess this choice empirically, we compared CatBoost with Random Forest and XGBoost using the same training-sample and feature framework. Predictive performance was measured using F1 score, and computational efficiency was represented by observed full-batch inference throughput (Sect. S1.2 and Table S2).”

B. Sect. 4.1.1, Lines 559–566 – We added the benchmark results and a calibrated interpretation of the predictive-performance and throughput differences.

“The classifier benchmark provided additional empirical support for selecting CatBoost (Table S2). CatBoost achieved the highest F1 score of 0.8358, compared with 0.832 for XGBoost and 0.828 for Random Forest. On the common NVIDIA A100 platform, CatBoost processed 3,148,493 samples/sec, compared with 719,361 samples/sec for XGBoost, corresponding to approximately 4.4 times the observed full-batch GPU inference throughput. Random Forest processed 163,614 samples/sec on the CPU platform. Because Random Forest and the two boosting models were evaluated on different processor architectures, the Random Forest throughput is reported as an implementation-specific reference rather than a hardware-normalized comparison. Overall, CatBoost provided the highest observed F1 score and the fastest GPU inference among the evaluated implementations.”

C. Sect. S1.2, Lines 62–75 – We added a supplementary subsection reporting the benchmark design, processing platforms, metrics, results, and hardware-comparison limitation.

“To evaluate the suitability of CatBoost for the high-dimensional global cropland-mapping workflow, we compared it with Random Forest and XGBoost using the same FAST-Crop plus augmented sample framework and the same 557-dimensional spectro-temporal feature representation. Predictive performance was summarized using Macro-F1, while computational efficiency was represented by observed full-batch inference throughput in samples per second.

XGBoost and CatBoost were evaluated on the same NVIDIA A100 GPU. Random Forest was evaluated on a dual-socket Intel Xeon Silver 4316 CPU

system with 40 physical cores and 80 logical CPUs. Because Random Forest and the two boosting models were executed on different processor architectures, their throughput values represent the observed implementations and should not be interpreted as hardware-normalized algorithmic comparisons. The XGBoost–CatBoost comparison, however, provides a direct same-platform comparison on the NVIDIA A100 GPU.

CatBoost achieved the highest Macro-F1 of 0.8358, compared with 0.8324 for XGBoost and 0.8283 for Random Forest (Table S2). CatBoost also achieved an inference throughput of 3,148,493 samples/sec, approximately 4.4 times the 719,361 samples/sec obtained by XGBoost on the same GPU platform. Random Forest processed 163,614 samples/sec on the CPU platform. These results indicate that CatBoost provided the best observed combination of predictive performance and GPU inference efficiency for the computationally intensive annual global mapping workflow."

D. Table S2, Lines 76 – We added the Macro-F1 score, observed full-batch inference throughput, and processing platform for each classifier.

Table S2. Predictive performance and observed full-batch inference throughput of the evaluated classifiers.

Classifier	F1 score	Inference throughput (samples/sec)	Processing platform
Random Forest	0.8283	163,614	Intel Xeon Silver 4316 CPU
XGBoost	0.8324	719,361	NVIDIA A100 GPU
CatBoost	0.8358	3,148,493	NVIDIA A100 GPU

Table note: XGBoost and CatBoost were evaluated on the same NVIDIA A100 GPU. Random Forest was evaluated on a CPU platform; its throughput is therefore reported as an implementation-specific reference rather than a hardware-normalized comparison.

(3) The independent validation appears to emphasize stable cropland, while transition areas are less well assessed. The manuscript states that the independent validation set was extracted from stable cropland records in FAST.

This supports overall map accuracy, but it does not fully test the product's ability to capture expansion, abandonment, active fallow, and other transitional states, which are central to the paper's concept of "structural evolution." A more targeted validation for transition pixels is needed.

> Thank you for highlighting this important limitation of the original validation. We agree that the stable-site FAST-Crop samples provide evidence of static Cropland/Non-cropland thematic accuracy but cannot directly assess crop gain, crop loss, transition timing, abandonment, or active-fallow dynamics. To address this concern, we added a targeted multi-temporal assessment using two external reference sources: the global GLAD Cropland samples, covering four transitions between adjacent four-year epochs, and repeated European LUCAS observations, covering five survey intervals from 2006 to 2022. GACED30 achieved higher recall and F1 than GLC_FCS30D for both crop gain and crop loss under both reference assessments. Nevertheless, the absolute transition scores remained modest, and GACED30 had slightly lower precision in the GLAD-based comparison. We therefore interpret the results as controlled comparative evidence of greater sensitivity to the evaluated transitions, not as validation of every annual layer or exact transition year. We also explicitly acknowledge that the available observations cannot distinguish permanent abandonment from active fallow, temporary interruption, or within-interval reversals.

The manuscript has been revised as follows:

A. Sect. 2.3.2, Lines 172–182 – We renamed the subsection 'Independent static validation samples' and clarified the stable-state scope of the FAST-Crop assessment.

"To provide an objective static thematic assessment of GACED30, we employed a validation set isolated entirely from model calibration. This subset was extracted from the temporally stable records of the FAST validation set (Li et al., 2017). The FAST project generated its validation data using a sampling

protocol distinct from that used for the training data: whereas training sites were manually selected for spectral representativeness, validation samples were allocated through a probabilistic global equal-area random sampling design. This probability-based framework provided spatially distributed observations of global land-cover heterogeneity. Unlike the augmented training samples, these validation samples were not subject to GLAD-based stratification or automated agreement filtering, thereby preventing GLAD-conditioned candidate selection from being propagated into this assessment. The final validation set comprised 29,226 sample locations, including 3,268 confirmed stable cropland samples. Because this subset deliberately emphasizes stable land-cover states, it was used to evaluate thematic agreement near the reference observation date but was not interpreted as direct validation of cropland gain, crop loss, or transition timing."

B. Sect. 2.3.3, Lines 184–196 – We added the GLAD Cropland and LUCAS multi-temporal reference samples and explained their complementary temporal and geographical coverage.

"To complement the static FAST-Crop assessment, we used two external multi-temporal reference sources whose reference labels were not used in GACED30 training or spatiotemporal post-processing. The first was the global probability-based reference sample associated with GLAD Cropland, comprising 3,500 sample locations with cropland interpretations for five four-year epochs: 2000–2003, 2004–2007, 2008–2011, 2012–2015, and 2016–2019 (Potapov et al., 2022). These observations supported the assessment of crop gain and crop loss between four pairs of adjacent epochs.

The second reference source was the European Union Land Use/Cover Area frame Survey (LUCAS), which provides harmonized in situ land-cover observations for the 2006, 2009, 2012, 2015, and 2018 survey campaigns (d'Andrimont et al., 2020), supplemented with observations from the 2022

campaign (d'Andrimont et al., 2024). Only locations with valid observations at both endpoints of a survey interval were retained.

GLAD Cropland provides global transition evidence across four adjacent four-year intervals, whereas LUCAS supplies denser repeated observations over five European survey intervals. Together, these datasets enable a targeted multi-temporal assessment of crop gain and crop loss, although their temporal spacing does not independently identify the exact year in which a transition occurred."

C. Fig. 2 and Sect. 3.5, Lines 401–407 – We reorganized the evaluation framework into five complementary components covering static thematic accuracy, multi-temporal transitions, regional and agro-ecological performance, area estimation and inter-product comparison, and national-statistics consistency.

"To assess the quality and reliability of GACED30, we adopted five complementary approaches (Fig. 2e): (1) static thematic accuracy assessment using the independent FAST-Crop validation samples; (2) multi-temporal assessment of crop gain and crop loss using GLAD Cropland and LUCAS reference samples; (3) matched regional evaluation in China together with agro-ecological-zone-stratified accuracy assessment; (4) global and regional area estimation and inter-comparison with existing global and regional cropland products; and (5) comparison with national agricultural statistics from FAOSTAT. These components respectively evaluate static thematic agreement, transition detection across multiple intervals, regional and environmental variations in performance, sample-adjusted area trajectories, and aggregated national-scale consistency."

D. Sect. 3.5.1, Lines 441–443 – We clarified that FAST-Crop evaluates static thematic agreement and is not direct validation of temporal transitions.

“Additionally, since the FAST-Crop validation subset predominantly represents stable land-cover states, these metrics evaluate static thematic agreement and are not interpreted as direct validation of cropland transitions.”

E. Sect. 3.5.2, Lines 445–470 – We added the complete transition-assessment procedure, including target-transition definitions, GLAD epoch aggregation, LUCAS survey pairing, class harmonization, and the selection of precision, recall, and F1 as the principal metrics.

“We evaluated two directional transitions separately: Cropland to Non-cropland, hereafter termed crop loss, and Non-cropland to Cropland, hereafter termed crop gain. For each directional assessment, the target transition was treated as the positive class, whereas stable pairs and the opposite transition were treated as negative cases. These sample-level transitions provide a targeted assessment of change detection but are not identical to the 1-km structural expansion and reduction classes, which are derived from trends fitted to the complete 25-year annual record.

For the GLAD Cropland assessment, the annual binary layers of GACED30 and GLC_FCS30D were aggregated within each corresponding four-year epoch using the pixel-wise median. Crop gain and crop loss were then determined for each of the four adjacent-epoch intervals: 2000–2003 to 2004–2007, 2004–2007 to 2008–2011, 2008–2011 to 2012–2015, and 2012–2015 to 2016–2019. The mapped transitions were evaluated against the corresponding GLAD reference transitions at the same sample locations. Each valid location–interval pair was treated as one observation, and the four intervals were pooled for the overall assessment. The same temporal aggregation, spatial extraction,

and class-harmonization procedures were applied to GACED30 and GLC_FCS30D.

For LUCAS, valid observations at the same location in successive available surveys were paired. The corresponding annual GACED30 and GLC_FCS30D classes were extracted for the two survey years, and the reference and mapped transitions were determined from their respective endpoint classes. When a location occurred in more than one survey interval, each location–interval pair was treated as one unweighted observation. Results from the 2006–2009, 2009–2012, 2012–2015, 2015–2018, and 2018–2022 intervals were pooled for the overall assessment.

Class harmonization was performed before determining the transitions. GACED30 class 10 was treated as Cropland. For LUCAS, land-cover codes B, B11–B19, B21–B23, B31–B37, B41–B45, B51–B54, B71–B77, B81–B84, BX1, and BX2 were mapped to Cropland. Temporary grassland (B55) was mapped to Grassland and was therefore excluded. For GLC_FCS30D, classes 10, 11, 12, and 20 were mapped to Cropland, whereas class 130 was treated as Grassland. Consequently, permanent crops were included, temporary grasslands were excluded, and fallow land was not automatically assigned to Cropland but followed its original observed or product land-cover class.

Because stable observations predominated in both reference sources, overall accuracy was not used as the principal transition metric, as it would be dominated by unchanged pairs. The resulting metrics were interpreted within the spatial, temporal, and semantic support of each reference dataset."

F. Sect. 4.1.2 and Table 3, Lines 577–598 – We added the quantitative crop-gain and crop-loss results for GACED30 and GLC_FCS30D against both reference sources, together with a qualified interpretation.

“The multi-temporal assessment (Table 3) provides a targeted evaluation of crop gain and crop loss that is not available from the stable-site FAST-Crop samples. Across the four pooled adjacent GLAD Cropland epoch intervals, GACED30 achieved higher recall and F1 score than GLC_FCS30D for both transition directions. For crop loss, its recall and F1 score were 0.177 and 0.178, compared with 0.129 and 0.152 for GLC_FCS30D. For crop gain, the corresponding values were 0.199 and 0.212 for GACED30 and 0.171 and 0.197 for GLC_FCS30D. GACED30 nevertheless had slightly lower precision: 0.180 versus 0.186 for crop loss and 0.227 versus 0.232 for crop gain.

The pooled LUCAS comparison showed larger differences between the products. For crop loss, GACED30 achieved precision, recall, and F1 score values of 0.099, 0.072, and 0.083, respectively, compared with 0.062, 0.016, and 0.026 for GLC_FCS30D. For crop gain, GACED30 obtained values of 0.113, 0.058, and 0.076, whereas GLC_FCS30D obtained 0.057, 0.014, and 0.023.

GACED30 therefore achieved higher transition recall and F1 score in both reference assessments, indicating greater sensitivity to the evaluated reference transitions. However, the absolute metrics remained modest, and its GLAD-based precision was slightly lower than that of GLC_FCS30D. Transition detection requires correct classification at both temporal endpoints, so an error at either endpoint can generate an incorrect change label. The low prevalence of transitions, temporal aggregation, spatial-support differences, and residual semantic differences further affect these metrics. The results provide a controlled common-sample comparison of transition detection across multiple intervals but do not establish definition-independent superiority or validate the exact year of annual changes.”

Table 3: Pooled transition detection by GACED30 and GLC_FCS30D using the GLAD Cropland and LUCAS multi-temporal reference samples

Reference set	Target transition	Product	Precision	Recall	F1
GLAD Cropland	Cropland → Non-cropland	GACED30	0.180	0.177	0.178
	Cropland → Non-cropland	GLC_FCS30D	0.186	0.129	0.152
	Non-cropland → Cropland	GACED30	0.227	0.199	0.212
	Non-cropland → Cropland	GLC_FCS30D	0.232	0.171	0.197
LUCAS	Cropland → Non-cropland	GACED30	0.099	0.072	0.083
	Cropland → Non-cropland	GLC_FCS30D	0.062	0.016	0.026
	Non-cropland → Cropland	GACED30	0.113	0.058	0.076
	Non-cropland → Cropland	GLC_FCS30D	0.057	0.014	0.023

G. Sects. 5.1 Lines 820–826 and 5.3 Lines **–** – We added a balanced interpretation of the transition results and explicitly stated that the available references cannot validate every annual layer, exact transition timing, or the causal meaning of crop loss.

“The complementary evaluations describe different aspects of product performance. The FAST-Crop assessment indicates strong static thematic agreement, the matched China assessment shows overall accuracy comparable with CACD, and the AEZ-stratified results identify geographically heterogeneous performance. The GLAD Cropland and LUCAS assessments provide direct evidence on crop-gain and crop-loss detection across multiple intervals: GACED30 achieved higher recall and F1 than GLC_FCS30D on the common samples, although the modest absolute scores demonstrate the continuing difficulty of transition mapping. Together, these results support the use of GACED30 for global and regional structural monitoring but do not constitute validation of the exact year or cause of every local change.”

“Moreover, the available GLAD Cropland and LUCAS references assess crop gain and crop loss across multi-year intervals rather than every annual layer. They cannot identify the exact transition year, detect all within-interval reversals,

or distinguish permanent cropland reduction from temporary non-cropland states. More independent annual and transition-specific reference observations are therefore needed"

H. Abstract Line 20-22 and Sect. 6 Line 896-898, We added concise statements reporting the complementary multi-temporal assessment while noting the modest absolute transition scores.

"*Multitemporal assessments using GLAD Cropland and LUCAS reference samples showed higher recall and F1 scores than GLC_FCS30D for both crop gain and crop loss, although the absolute transition scores remained modest.*"

"*Multitemporal assessments yielded higher recall and F1 scores than GLC_FCS30D for both crop gain and crop loss; however, the modest absolute transition scores indicate that local annual changes should be interpreted cautiously*"

(4) The discussion should expand the practical applications of the dataset. The manuscript already mentions implications for global sustainability, telecoupling, carbon accounting, and Earth system modelling, but these applications are still described rather generally. This section would be stronger if the authors more explicitly discussed how the dataset could be used in food security assessment, agricultural frontier monitoring, land-use policy evaluation, carbon accounting, and Earth system model benchmarking.

> Thank you for this constructive suggestion. We agree that the original discussion named several potential applications without explaining how GACED30 would contribute to them or what complementary information would be required. We therefore revised Sect. 5.2 to provide concrete and appropriately qualified application pathways. The revised text explains that GACED30 can supply the cropland-extent component of food-security assessments; identify candidate agricultural-frontier and persistent cropland-

reduction zones; support telecoupled analyses when combined with trade-flow and socioeconomic information; serve as a spatial outcome layer in appropriately designed policy evaluations; provide activity information for carbon-accounting analyses when combined with preceding land cover, carbon stocks, and emission factors; and support comparison with land-use inputs or outputs used in Earth system models. We also clarify that GACED30 alone does not measure food security, quantify carbon emissions, establish policy effects, or identify the causal drivers of mapped changes.

The manuscript has been revised as follows:

A. Sect. 5.2, Lines 833–847 – We expanded the first paragraph to cover food-security assessment, agricultural-frontier monitoring, telecoupled land-system analysis, and policy evaluation, while the second paragraph now provides concrete pathways for carbon accounting and Earth system model benchmarking. The following sentence is inserted immediately after the sentence ending with '…land-tenure information' and before the paragraph beginning 'The structural trend record also has practical value…':

"The results in Sects. 4.3.1–4.3.3 indicate that global cropland increased modestly from 1458.5 Mha in 2000 to 1488.5 Mha in 2024, while stable cropland remained predominant and statistically significant expansion and reduction were spatially concentrated. This combination shows that a relatively small global net change can coexist with substantial regional reorganisation, particularly the contrast between concentrated expansion in parts of Africa and South America and widespread stability or reduction in much of the Global North. Because GACED30 maps cropland extent rather than crop type, yield, production, or food access, it cannot measure food security directly. When combined with crop-type maps, yield and production statistics, population, market-access, and nutrition data, however, it can provide the land-footprint component of food-security assessments and help distinguish cases in which production changes are accompanied by cropland expansion from those

occurring without marked extent change. The significant trend layer can also identify candidate agricultural-frontier and persistent cropland-reduction zones for targeted monitoring; determining the preceding land cover and environmental implications requires independent forest, grassland, biodiversity, protected-area, infrastructure, and land-tenure information. When combined with trade-flow and socioeconomic data, these spatial trajectories can support telecoupled land-system analyses linking distant production and consumption regions; when paired with policy boundaries, implementation dates, and an appropriate comparison or counterfactual design, they can also serve as an outcome layer for land-use policy evaluation, although GACED30 alone cannot establish causal effects."

B. Sect. 5.3, Lines 886–889 – We clarified the additional evidence and analytical designs required for driver attribution and policy evaluation.

"Third, the Structural Evolution Indicators should be combined with independent environmental, socioeconomic, policy, and land-cover data in explicitly designed attribution or counterfactual analyses. Such analyses are required before the regional associations identified here can be interpreted as causal drivers of cropland change."

(5) The product is explicitly aligned with the FAO definition and includes permanent woody crops, active fallow, and agricultural structures, while excluding temporary meadows and pastures. This is reasonable, but it also means that part of the disagreement with other global products may be driven by definitional differences rather than mapping quality alone. The manuscript should make this point more explicit in both the validation and discussion sections.

> Thank you for this helpful comment. We agree that disagreement among products should not be interpreted exclusively as mapping error. GACED30

includes orchards and other permanent woody crops and active fallow within Cropland while excluding temporary meadows and pastures, whereas the native legends of the evaluated products treat these categories differently. Consequently, even under a common reference-sample assessment, part of the observed omission or commission disagreement can reflect differences in semantic scope. We have now made this qualification explicit in the validation methodology, accuracy results, global area comparison, and discussion, and we have revised the interpretation so that the reported comparisons indicate agreement with the adopted FAST-Crop/GACED30 definition rather than universal superiority across all cropland definitions.

The manuscript has been revised as follows:

A. Sect. 3.5.1, Lines 438–441 – We added a direct qualification to the common-reference comparison.

“All products were evaluated against the same FAST-Crop reference labels following the cropland definition described in Sect. 2.1. Because the native product legends differ in their treatment of orchards and other perennial woody crops, active fallow, and temporary meadows and pastures, part of the observed disagreement may reflect definitional differences.”

B. Sect. 4.1.1, Lines 552–558 – We clarified how cropland definitions and temporal mismatches affect interpretation of the accuracy comparison.

“Part of the mismatch with peer products may reflect differences in cropland definitions rather than mapping errors alone. Specifically, GACED30 includes orchards and active fallow but excludes temporary meadows and pastures, whereas peer products differ in their treatment of these categories. Temporal mismatches between the validation samples and product layers may also contribute to the observed differences. The metrics in Table 2 should therefore be interpreted as the products' relative agreement with the adopted FAST-

Crop/GACED30 definition rather than as evidence of universal superiority across all cropland categories. Accordingly, Table 2 supports the static thematic reliability of GACED30 under the adopted Cropland/Non-cropland reference definition but does not directly assess its ability to detect inter-period cropland changes."

C. Sect. 4.3.1, Lines 710–719 – We replaced the previous attribution to a single 'Permanent Crop Gap' with an interpretation based on product-specific mapping behaviour and native definitions.

"The consistently adjusted peer-product series showed different area levels and endpoint changes over their respective periods. GLC_FCS30D increased from 1347.2 Mha in 2000 to 1401.8 Mha in 2022, whereas GLAD Cropland increased from 1424.6 Mha for the 2000–2003 epoch to 1514.5 Mha for the 2016–2019 epoch. Over its shorter temporal coverage, Esri Land Cover increased from 1511.1 Mha in 2017 to 1561.4 Mha in 2024. ESA WorldCover showed little change between its two available years, decreasing from 1520.8 Mha in 2020 to 1516.3 Mha in 2021. These results show that the adjustment does not force the products toward a common trajectory; substantial differences in both estimated area and temporal evolution remain. GACED30 provides a continuous 25-year annual record with a comparatively gradual long-term increase. Across products, the adjusted estimates form a continuum shaped by product-specific mapping behavior and native definitions, particularly their treatment of perennial woody crops, active fallow, temporary meadows and pastures, and mixed agricultural mosaics."

D. Sect. 5.1, Lines 827–831 – We clarified that the adjusted inter-product trajectories must be interpreted in relation to product definitions, mapping strategies, and the adopted structural-trend criteria.

“Differences among the adjusted inter-product trajectories should also be interpreted in relation to product definitions and mapping strategies. GACED30 includes permanent woody crops and active fallow while excluding temporary meadows and pastures, and its GLAD-conditioned augmentation changes the representation of individual cropland subclasses. Consequently, the comparatively gradual global increase and the associated regional patterns should be understood within the adopted cropland definition and statistical trend criteria rather than as evidence that alternative products are erroneous.”

Specific Comments

(1) It would help to report performance separately by region or agro-ecological zone, rather than only globally.

> Thank you for this valuable suggestion. We agree that globally aggregated metrics can conceal substantial geographical variation. We therefore added an assessment stratified by 18 thermal-moisture agro-ecological zones derived from GAEZ v4. Across the tropical, subtropical, and temperate strata represented by AEZs 1–15, overall accuracy ranged from 0.931 to 0.997 and F1 ranged from 0.631 to 0.913. The cool, humid tropical stratum had a producer’s accuracy of only 0.498 despite an overall accuracy of 0.951, showing that high overall accuracy can conceal substantial cropland omission in a class-imbalanced environment. The boreal sub-humid and humid strata contained no positive Cropland reference cases, so their class-specific metrics are reported as not applicable. These results are now presented in the main text, the Supplement, and the limitations discussion.

The manuscript has been revised as follows:

A. Sect. 2.5, Lines 223–227 – We introduced GAEZ v4 and the 18 thermal-moisture agro-ecological strata used for geographical performance diagnosis.

“To examine geographical variation in mapping performance, we used the Global Agro-Ecological Zones version 4 dataset (GAEZ v4; FAO and IIASA,

2021). The GAEZ thermal and moisture regimes were organized into 18 broad analytical strata spanning tropical, subtropical, temperate, and boreal environments under semi-arid, sub-humid, and humid moisture conditions. These strata were used solely to diagnose geographical variations in accuracy rather than to evaluate agricultural suitability."

B. Sect. 3.5.3, Lines 486–489 – We added the procedure used to assign FAST-Crop validation records to agro-ecological strata and calculate stratum-specific metrics.

"Finally, each independent FAST-Crop static validation record was spatially assigned to one of the 18 agro-ecological strata derived from GAEZ v4. OA, UA, PA, F1 score, and Kappa were then calculated separately for each stratum. Where a stratum contained no positive Cropland reference cases, cropland-specific UA, PA, F1, and Kappa were considered not applicable and reported as NA rather than as zero accuracy."

C. Sect. 4.1.3, Lines 615–622 – We added the principal AEZ-stratified findings and highlighted the lower producer's accuracy in the cool, humid tropical stratum.

"The AEZ-stratified results further showed that global accuracy was not spatially uniform (Table S4). Across the tropical, subtropical, and temperate strata represented by AEZs 1–15, OA ranged from 0.931 to 0.997 and F1 ranged from 0.631 to 0.913. The warm, humid subtropical stratum achieved the highest F1 score among these zones (0.913), with a UA of 0.905 and a PA of 0.922. The lowest F1 occurred in the cool, humid tropical stratum (0.631), where the relatively high UA of 0.861 was accompanied by a substantially lower PA of 0.498. This contrast indicates that omission of cropland was the principal source of error in this environment and that its high OA of 0.951 partly reflected

the predominance of Non-cropland reference records. The boreal sub-humid and humid strata contained no positive Cropland reference cases and therefore did not support the calculation of cropland-specific accuracy metrics."

D. Table S4, Supplement, Lines 120– We added the complete accuracy results for all 18 agro-ecological strata and report class-specific metrics as NA where no positive Cropland references were available.

Table S4. Static thematic accuracy of GACED30 stratified by 18 agro-ecological zones.

AEZ	Agro-ecological zone	OA	UA	PA	F1	Kappa
1	Tropics, warm, semi-arid	0.992	0.914	0.756	0.827	0.823
2	Tropics, warm, sub-humid	0.964	0.879	0.928	0.903	0.881
3	Tropics, warm, humid	0.953	0.863	0.946	0.902	0.871
4	Tropics, cool, semi-arid	0.951	0.896	0.824	0.859	0.829
5	Tropics, cool, sub-humid	0.951	0.868	0.754	0.807	0.779
6	Tropics, cool, humid	0.951	0.861	0.498	0.631	0.607
7	Subtropics, warm, semi-arid	0.989	0.895	0.804	0.847	0.841
8	Subtropics, warm, sub-humid	0.949	0.897	0.897	0.897	0.863
9	Subtropics, warm, humid	0.953	0.905	0.922	0.913	0.881
10	Subtropics, cool, semi-arid	0.931	0.888	0.874	0.881	0.832
11	Subtropics, cool, sub-humid	0.941	0.888	0.889	0.888	0.848
12	Subtropics, cool, humid	0.938	0.845	0.743	0.790	0.754
13	Temperate, semi-arid	0.987	0.896	0.812	0.852	0.845
14	Temperate, sub-humid	0.997	0.984	0.759	0.857	0.855
15	Temperate, humid	0.994	0.900	0.738	0.811	0.808
16	Boreal, semi-arid	1.000	1.000	1.000	1.000	1.000
17	Boreal, sub-humid	\	\	\	\	\
18	Boreal, humid	\	\	\	\	\

E. Sect. 5.3, Lines 859–867 – We integrated the AEZ results into the limitations discussion and identified environments requiring additional regional validation.

“The principal spatial and semantic limitations of GACED30 arise from its binary class structure and 30-m spatial resolution. Active cultivation and rotational fallow are represented within a single Cropland class, so the dataset does not distinguish their annual management status. Fields smaller than or comparable to a Landsat pixel may be omitted or spatially smoothed, particularly in fragmented smallholder landscapes. Temporary fallow, ploughed pasture, degraded rangeland, and natural bare or sparsely vegetated surfaces can also exhibit similar spectral and phenological characteristics. The AEZ assessment illustrates this geographical heterogeneity: in the cool, humid tropical stratum, the producer’s accuracy of 0.498 indicates substantial cropland omission despite a high overall accuracy. The boreal sub-humid and humid reference subsets contained no positive Cropland cases and therefore did not support class-specific evaluation. Local applications in these environments require additional regional validation.”

(2) The discussion of potential applications should be more concrete. For example, please elaborate on use cases in food security, agricultural expansion monitoring, telecoupled land-system analysis, carbon accounting, and policy evaluation.

> Thank you for this suggestion. As detailed in our response to Major Comment 4, we substantially revised Sect. 5.2 to connect the demonstrated properties of GACED30 with concrete applications in food-security assessment, agricultural-frontier monitoring, telecoupled land-system analysis, carbon accounting, Earth system model benchmarking, and policy evaluation. For each application, we now identify the complementary environmental, agricultural, socioeconomic, trade, or policy information required and clarify that GACED30 provides a cropland-extent or spatial-outcome layer rather than direct evidence of food security, emissions, policy effects, or causal drivers.

The manuscript has been revised as follows:

A. Sect. 5.2, Lines 833–847 – We expanded the first paragraph to cover food-security assessment, agricultural-frontier monitoring, telecoupled land-system analysis, and policy evaluation, while the second paragraph now provides concrete pathways for carbon accounting and Earth system model benchmarking. The following sentence is inserted immediately after the sentence ending with '…land-tenure information' and before the paragraph beginning 'The structural trend record also has practical value…':

"The results in Sects. 4.3.1–4.3.3 indicate that global cropland increased modestly from 1458.5 Mha in 2000 to 1488.5 Mha in 2024, while stable cropland remained predominant and statistically significant expansion and reduction were spatially concentrated. This combination shows that a relatively small global net change can coexist with substantial regional reorganisation, particularly the contrast between concentrated expansion in parts of Africa and South America and widespread stability or reduction in much of the Global North. Because GACED30 maps cropland extent rather than crop type, yield, production, or food access, it cannot measure food security directly. When combined with crop-type maps, yield and production statistics, population, market-access, and nutrition data, however, it can provide the land-footprint component of food-security assessments and help distinguish cases in which production changes are accompanied by cropland expansion from those occurring without marked extent change. The significant trend layer can also identify candidate agricultural-frontier and persistent cropland-reduction zones for targeted monitoring; determining the preceding land cover and environmental implications requires independent forest, grassland, biodiversity, protected-area, infrastructure, and land-tenure information. When combined with trade-flow and socioeconomic data, these spatial trajectories can support telecoupled land-system analyses linking distant production and consumption regions; when paired with policy boundaries, implementation dates, and an

appropriate comparison or counterfactual design, they can also serve as an outcome layer for land-use policy evaluation, although GACED30 alone cannot establish causal effects."

B. Sect. 5.3, Lines 886–889 – We clarified the additional evidence and analytical designs required for driver attribution and policy evaluation.

"Third, the Structural Evolution Indicators should be combined with independent environmental, socioeconomic, policy, and land-cover data in explicitly designed attribution or counterfactual analyses. Such analyses are required before the regional associations identified here can be interpreted as causal drivers of cropland change."

(3) Please clarify more explicitly that part of the mismatch with peer products may arise from different cropland definitions, especially regarding orchards, active fallow, and temporary meadows/pastures.

> Thank you for pointing this out. We agree that the observed mismatch with peer products should not be attributed solely to mapping error. As detailed in our response to Major Comment 5, the validation methodology and discussion now explicitly explain that GACED30 includes orchards and active fallow but excludes temporary meadows and pastures, whereas peer-product definitions differ. The accuracy results have therefore been qualified as measuring relative agreement with the adopted FAST-Crop/GACED30 definition rather than demonstrating definition-independent superiority.

The manuscript has been revised as follows:

A. Sect. 4.1.1, Lines 552–558 – We clarified how cropland definitions and temporal mismatches affect interpretation of the accuracy comparison.

"Part of the mismatch with peer products may reflect differences in cropland definitions rather than mapping errors alone. Specifically, GACED30 includes

orchards and active fallow but excludes temporary meadows and pastures, whereas peer products differ in their treatment of these categories. Temporal mismatches between the validation samples and product layers may also contribute to the observed differences. The metrics in Table 2 should therefore be interpreted as the products' relative agreement with the adopted FAST-Crop/GACED30 definition rather than as evidence of universal superiority across all cropland categories. Accordingly, Table 2 supports the static thematic reliability of GACED30 under the adopted Cropland/Non-cropland reference definition but does not directly assess its ability to detect inter-period cropland changes."

(4) The manuscript would benefit from a short paragraph stating where the dataset is expected to perform less well, for example in smallholder mosaics, highly fragmented systems, or regions with strong semantic ambiguity between cropland and grassland.

> Thank you for this helpful suggestion. We agree that the manuscript should clearly identify the landscape and semantic conditions under which GACED30 is less reliable. We therefore revised Sect. 5.3 to explain that fields smaller than or comparable to a Landsat pixel may be omitted or spatially smoothed, particularly in fragmented smallholder landscapes, and that uncertainty increases where temporary fallow, ploughed pasture, degraded rangeland, and bare or sparsely vegetated surfaces have similar spectral and phenological characteristics. We also connected these expected limitations to the AEZ assessment, which identified substantial cropland omission in the cool, humid tropical stratum, and now recommend additional regional validation for local applications in these environments.

The manuscript has been revised as follows:

A. Sect. 5.3, Lines 859–867 – We added a concise, evidence-based account of the spatial and semantic settings in which GACED30 is expected to perform less well.

“The principal spatial and semantic limitations of GACED30 arise from its binary class structure and 30-m spatial resolution. Active cultivation and rotational fallow are represented within a single Cropland class, so the dataset does not distinguish their annual management status. Fields smaller than or comparable to a Landsat pixel may be omitted or spatially smoothed, particularly in fragmented smallholder landscapes. Temporary fallow, ploughed pasture, degraded rangeland, and natural bare or sparsely vegetated surfaces can also exhibit similar spectral and phenological characteristics. The AEZ assessment illustrates this geographical heterogeneity: in the cool, humid tropical stratum, the producer’s accuracy of 0.498 indicates substantial cropland omission despite a high overall accuracy. The boreal sub-humid and humid reference subsets contained no positive Cropland cases and therefore did not support class-specific evaluation. Local applications in these environments require additional regional validation.”

(5) Please check the data information. The pixel values in each band indicate land-cover classification status; however, I get quick look at the data, and it seems that cropland is coded as 10.

> Thank you for checking the data information and drawing our attention to this ambiguity. You are correct that cropland is intentionally encoded as 10 in each annual GACED30 raster layer, following the FROM-GLC coding convention, while non-cropland is encoded as 0. The value 10 therefore denotes the binary Cropland class and is not an error in the raster data. We have added an explicit coding statement to the Data Availability section and corrected the Zenodo dataset description so that users can interpret the pixel values unambiguously.

The manuscript has been revised as follows:

A. Data Availability, Lines 906–907 – We added an explicit statement defining the pixel codes used in every annual raster layer.

“Each annual raster layer follows the FROM-GLC coding convention, in which a value of 10 denotes cropland and a value of 0 denotes non-cropland.”

B. Zenodo dataset description – We corrected the online data description to state explicitly that the cropland class is coded as 10 and the non-cropland class as 0.

Comments by Reviewer #3

General comments

This study presents the annual updated cropland dataset at 30m resolution for the year 2000-2024. The proposed method integrates gap-free SDC30 time series with spectral-semantic sample alignment to generate such valuable dataset. Such datasets are essential for both crop monitoring community and land use land cover domains as it makes a substantial contribution to the Earth system science community by providing a temporally coherent, high-resolution dataset. While the manuscript is potentially publishable, there are still several major concerns, in particular the mixture of the different sub-classes of cropland. The samples used for training is binary cropland and non-cropland while the intra-class variations of different croplands (permanent crops, temporary crops, active fallow, etc) will hamper the classification accuracy. Meanwhile, in its current form, the manuscript suffers from significant shortcomings in methodological transparency, validation design, and demonstration of novelty relative to existing products. These deficiencies must be addressed before the dataset and its documentation meet the publication standards of Earth System Science Data.

> Thank you for your constructive and comprehensive assessment. We appreciate your recognition of the value of a global annual 30-m cropland dataset, while agreeing that the submitted manuscript required greater methodological transparency, stronger validation of cropland transitions, and more cautious interpretation of intra-class performance and interannual consistency. In response, we have explicitly documented the dependence of the augmented training samples on the GLAD Cropland prior, quantified the resulting change in cropland-subclass composition, and added a controlled training-set ablation analysis. We have reframed GACED30 as a harmonized annual baseline rather than a fully temporally constrained sequence, expanded the validation framework to include multi-temporal crop-gain and crop-loss

assessments using GLAD Cropland and LUCAS reference samples, applied a common sample-based area-adjustment framework to all compared products, and clarified the limitations of distinguishing active fallow, pasture, rangeland, and bare surfaces at 30 m. We have also moderated claims of product superiority, added threshold-sensitivity analysis and agro-ecological diagnostics, corrected the affected figures and table entry, and streamlined the Data Availability section. We believe these revisions substantially improve the transparency, evidential support, and appropriate positioning of the dataset.

Major Comments

(1) The manuscript describes a spectral-semantic sample alignment strategy where candidate locations are stratified using the GLAD Cropland product and subsequently filtered by an expert-trained model. The final training set is, by design, heavily conditioned on the spatial priors of GLAD. This could artificially suppress the detection of cropland types that GLAD systematically misses (e.g., tree crops, permanent crops, etc). The authors must provide a rigorous analysis of the independence of the final training sample set from the GLAD prior. Specifically, what is the proportion of "augmented" samples that fell outside of GLAD cropland masks? Without this, the claim of outperforming GLAD in areas where GLAD is known to be weak is not fully substantiated. At the same time, I suggest to consider the dataset described in the peer-reviewed paper ‘A global reference dataset for land cover mapping at 10 m resolution’ to amend the training set.

> Thank you for identifying this important dependence. We agree that the original manuscript did not sufficiently disclose how the final training set remained influenced by the GLAD Cropland spatial prior, particularly for cropland categories that GLAD does not consistently represent, such as permanent woody crops and agricultural structures.

For the literal question concerning only the spatially augmented Cropland samples, none fell outside the GLAD Cropland mask by construction: the

augmented candidates were stratified using GLAD and retained only when the FAST-Crop-trained classifier agreed with the prior label. The augmented Cropland tier therefore cannot independently recover cropland omitted by GLAD. We agree, however, that the more informative measure of the final training set's independence is the representation contributed by the expert-annotated FAST-Crop tier. Orchard and Greenhouse samples, which represent categories not consistently included as cropland by GLAD, accounted for approximately 11% of the FAST-Crop Cropland subset. After the GLAD-conditioned augmented samples were added, these categories accounted for approximately 3% of the final combined Cropland training set. We now report this compositional change explicitly and acknowledge that augmentation increased geographical coverage while reducing the relative representation of categories outside the GLAD definition.

To evaluate whether this compositional shift affected binary recognition of the individual FAST-Crop subclasses, we added a controlled training-set ablation experiment. Adding the augmented samples increased F1 by 0.025 for Other Farmland, by 0.035 for Bare Farmland, and by 0.055 for Greenhouse. Orchard showed a smaller decrease of 0.017, from 0.574 to 0.557, while Rice Paddy decreased by 0.055. These results do not demonstrate independence from GLAD, but they show that the FAST-Crop tier continued to provide information for orchard recognition after augmentation. We therefore interpret the ablation as partly qualifying, rather than eliminating, the potential limitation caused by the reduced representation of GLAD-excluded cropland categories.

We also agree that the original comparison with GLAD Cropland was too strong. We have revised the Puglia example and the general accuracy interpretation so that broader mapped coverage is described as consistent with the GACED30 definition rather than as quantitative evidence of superior orchard detection. Finally, we carefully considered your suggestion to use the global 10-m reference dataset of Lesiv et al. (2025). Although it provides

valuable independent reference locations, its Crops class primarily represents annually harvested crops, while perennial woody crops may be assigned to Trees or Shrubs and its Fallow/Shifting Cultivation class is not directly equivalent to the GACED30 active-fallow definition. Directly merging its original labels would therefore not necessarily resolve the same semantic under-representation. We have identified harmonized reinterpretation of these samples as a priority for future training-set development.

The manuscript has been revised as follows:

A. Abstract and Sect. 1, Lines 16–22 – We moderated the description of sample alignment and the inter-product performance claims, and we now distinguish strong agreement under the adopted reference definition from definition-independent superiority.

“The mapping framework combines gap-free SDC30 observations, spectral-semantic alignment of expert-annotated and spatially augmented samples, and rule-based spatial and temporal refinement. Independent classifiers were trained for each year using a common observation, feature, sample-construction, and modeling protocol, and the resulting annual record was used to derive pixel-level Cropping Frequency and 1 km Structural Evolution Indicators. Against independent stable-site FAST-Crop samples, GACED30 achieved an overall accuracy of 96.5% and an F1 score of 0.844. Multitemporal assessments using GLAD Cropland and LUCAS reference samples showed higher recall and F1 scores than GLC_FCS30D for both crop gain and crop loss, although the absolute transition scores remained modest.”

B. Sect. 2.2, Lines 142–148 – We clarified the role and definitional scope of the GLAD Cropland prior and explained that it was used for spatial stratification rather than as a complete representation of the GACED30 definition.

“Further, the GLAD Cropland product served as a thematic stratification layer for generating the spatially augmented candidate pool. Its long-term record (Potapov et al., 2022) enabled candidate sampling across established cultivation zones and under-represented ecoregions. However, GLAD principally represents annual and perennial herbaceous crops and excludes tree crops, whereas GACED30 additionally includes permanent woody crops and agricultural structures. GLAD was therefore used as a spatial prior rather than as a complete representation of the GACED30 cropland definition, and its candidate labels were subsequently screened using a classifier trained on the expert-annotated FAST-Crop samples. The GLAD Cropland product and its associated validation sample set are available from https://glad.umd.edu/dataset/croplands.”

C. Sects. 2.3.1 and 3.2.1, Lines 158–170 – We quantified the representation of Orchard and Greenhouse samples, reported their reduction from approximately 11% of the FAST-Crop Cropland subset to approximately 3% of the final Cropland training set, and explicitly acknowledged that the augmented samples remained dependent on the GLAD spatial prior.

“Within the FAST-Crop cropland subset, Orchard and Greenhouse samples, which represent categories not consistently included as cropland by GLAD, accounted for approximately 11% of the samples.

To expand the geographical coverage of this expert-annotated baseline, we generated a spatially augmented sample set using global equal-area sampling and GLAD-based thematic stratification, followed by agreement screening with a FAST-Crop-trained classifier, spatial-spectral distribution screening,

performance-guided subset selection, and year-specific weak-label screening (Sect. 3.2.1 and Sects. S1.1–S1.2). This multi-stage procedure yielded 332,471 retained sample locations. Because the augmented cropland tier predominantly represented categories included within the GLAD cropland definition, adding these samples reduced the combined proportion of Orchard and Greenhouse samples from approximately 11% in the FAST-Crop cropland subset to approximately 3% in the final cropland training set. The potential effect of this compositional shift was evaluated through the training-set ablation experiment reported in Sect. 4.1.1 and Sect. S1.3."

D. Sect. S1.3 and Table S3, Lines 80–101 – We added a controlled training-set ablation experiment comparing FAST-Crop-only training with the final combined training set and reported subclass-specific binary macro-F1 scores.

Table S3. Subclass-specific binary macro-F1 scores for classifiers trained with the FAST-Crop samples alone and with the final combined training set.

FAST-Crop subclass	FAST-Crop only	FAST-Crop + augmented	Change in macro-F1
Rice Paddy	0.504	0.449	-0.055
Greenhouse	0.400	0.455	+0.055
Other Farmland	0.433	0.458	+0.025
Orchard	0.574	0.557	-0.017
Bare Farmland	0.464	0.498	+0.035

Note: Each row represents a separate binary comparison between the indicated cropland subclass and the non-cropland evaluation samples. The values do not constitute a multiclass assessment of confusion among cropland subclasses. The Greenhouse result should be interpreted cautiously.

"To assess whether the change in training-sample composition affected the binary recognition of individual FAST-Crop subclasses, we performed a controlled training-set ablation experiment. Two CatBoost classifiers were trained using the same modelling protocol. The first classifier used only the expert-annotated FAST-Crop training samples, whereas the second used the

final training set combining the FAST-Crop and filtered spatially augmented samples. Both classifiers were evaluated using the same subclass-labelled FAST-Crop evaluation samples.

Because the GACED30 classifier produces binary Cropland and Non-cropland predictions, this experiment was not a multiclass subclass assessment. Instead, each cropland subclass was evaluated separately against the non-cropland evaluation samples, with other cropland subclasses excluded from that comparison. The resulting binary macro-F1 scores therefore measure whether samples belonging to a given subclass continued to be distinguished from non-cropland after augmentation. They do not quantify confusion among the individual cropland subclasses.

Adding the spatially augmented samples increased macro-F1 from 0.433 to 0.458 for Other Farmland and from 0.464 to 0.498 for Bare Farmland. Orchard showed a smaller numerical decrease from 0.574 to 0.557, whereas Rice Paddy decreased from 0.504 to 0.449. The Greenhouse score increased from 0.400 to 0.455, but only six Greenhouse evaluation samples were available, precluding a reliable interpretation. These descriptive results indicate that augmentation improved binary performance for the two largest conventional subclasses while producing a small observed decrease for Orchard. The FAST-Crop tier therefore continued to provide the classifier with information for recognizing orchards as cropland, although the analysis does not eliminate the potential limitation associated with their reduced representation in the combined training set."

E. Sect. 4.1.1, Lines **–** – We qualified the inter-product accuracy comparison and summarized the subclass-ablation results without treating them as a multiclass assessment or proof of independence from GLAD.

“Part of the mismatch with peer products may reflect differences in cropland definitions rather than mapping errors alone. Specifically, GACED30 includes orchards and active fallow but excludes temporary meadows and pastures, whereas peer products differ in their treatment of these categories. Temporal mismatches between the validation samples and product layers may also contribute to the observed differences. The metrics in Table 2 should therefore be interpreted as the products' relative agreement with the adopted FAST-Crop/GACED30 definition rather than as evidence of universal superiority across all cropland categories. Accordingly, Table 2 supports the static thematic reliability of GACED30 under the adopted Cropland/Non-cropland reference definition but does not directly assess its ability to detect inter-period cropland changes.

The classifier benchmark provided additional empirical support for selecting CatBoost (Table S2). CatBoost achieved the highest F1 score of 0.8358, compared with 0.832 for XGBoost and 0.828 for Random Forest. On the common NVIDIA A100 platform, CatBoost processed 3,148,493 samples/sec, compared with 719,361 samples/sec for XGBoost, corresponding to approximately 4.4 times the observed full-batch GPU inference throughput. Random Forest processed 163,614 samples/sec on the CPU platform. Because Random Forest and the two boosting models were evaluated on different processor architectures, the Random Forest throughput is reported as an implementation-specific reference rather than a hardware-normalized comparison. Overall, CatBoost provided the highest observed F1 score and the fastest GPU inference among the evaluated implementations.

We further conducted a training-set ablation experiment to examine whether adding the GLAD-conditioned augmented samples affected binary

cropland recognition for the second-level FAST-Crop subclasses (Table S3). Adding the augmented samples increased F1 score by 0.025 for Other Farmland, by 0.035 for Bare Farmland and by 0.055 for the Greenhouse. Orchard showed a smaller decrease of 0.017, from 0.574 to 0.557, whereas Rice Paddy decreased by 0.055. Thus, despite the reduction in the relative representation of Orchard and Greenhouse samples, the FAST-Crop tier continued to support orchard recognition, with only a small observed decrease in Orchard performance. Nevertheless, the analysis qualifies rather than eliminates the potential influence of GLAD-conditioned augmentation on categories outside the GLAD cropland definition. Because each subclass was evaluated separately against non-cropland, these results should be interpreted as a binary sensitivity analysis rather than a multiclass subclass assessment."

F. Sect. 4.2.1, Lines 665–680 – We revised the Puglia comparison and the other qualitative examples so that they illustrate differences in mapped spatial patterns rather than serve as case-specific validation of particular cropland subclasses.

"To illustrate inter-product differences in challenging agricultural landscapes, we compared the 2020 GACED30 layer with temporally corresponding GLC_FCS30D and ESA WorldCover layers and the 2019 GLAD Cropland epoch (Fig. 4). The selected cases include cloud-prone, semi-arid, fragmented, and spectrally ambiguous agricultural environments and are intended to illustrate differences in mapped spatial patterns rather than provide case-specific accuracy validation.

In the Willamette Valley, USA (Fig. 4a), the products differ in their representation of managed grass-seed fields that may be spectrally similar to pasture and hay (Mueller-Warrant et al., 2011). Substantial differences are also visible in the drought-prone Murray–Darling Basin, Australia (Fig. 4d), where fallow cropland, pasture, and semi-arid vegetation can be difficult to distinguish

(Xie et al., 2024). In Puglia, Italy (Fig. 4b), GACED30 maps broader cropland coverage across the mixed olive-grove and annual-crop landscape than GLAD Cropland, consistent with the inclusion of permanent woody crops in the GACED30 definition (Damianidis et al., 2021). In the Gezira Scheme, Sudan (Fig. 4c), the products differ in their treatment of temporarily bare fields within the irrigated agricultural landscape.

Differences are also apparent in the fragmented paddy and aquaculture landscape of the Ganges Delta, Bangladesh (Fig. 4e), and in the persistently cloud-prone Chocó Department, Colombia (Fig. 4f), where GACED30 produces a comparatively continuous mapped cropland pattern. These examples demonstrate how product definitions, observation strategies, and temporal representations influence mapped cropland patterns; without independent case-specific reference data, they should not be interpreted as quantitative evidence that one product is universally more accurate."

G. Sects. 5.1 Lines 827-831 and 5.3, Lines 876–881 – We added the GLAD-conditioned sample composition as an explicit methodological trade-off and limitation, and clarified that strong binary accuracy does not imply equally well-constrained performance for every cropland subtype.

"Differences among the adjusted inter-product trajectories should also be interpreted in relation to product definitions and mapping strategies. GACED30 includes permanent woody crops and active fallow while excluding temporary meadows and pastures, and its GLAD-conditioned augmentation changes the representation of individual cropland subclasses. Consequently, the comparatively gradual global increase and the associated regional patterns should be understood within the adopted cropland definition and statistical trend criteria rather than as evidence that alternative products are erroneous."

"*The augmented Cropland layer also retains the influence of the GLAD Cropland spatial prior. Because GLAD Cropland excludes some categories included in GACED30, permanent woody crops and agricultural structures were represented primarily by the FAST-Crop tier and accounted for approximately 3% of the final combined Cropland training set. The ablation experiment indicated that the effects of augmentation varied among FAST-Crop subclasses, including a small decrease in Orchard performance. Strong binary cropland accuracy should therefore not be interpreted as demonstrating equally well-constrained performance for every cropland subtype.*"

H. Sect. 5.3 and References, Lines **–** – We added Lesiv et al. (2025) and explained that its reference locations require reinterpretation and harmonization with the GACED30 definition before they can be incorporated into future training-set development.

"Future development should prioritize three directions. First, integration of 10-m observations and in-season phenological information could improve the representation of small fields and enable more detailed characterization of annual management. Second, additional independent reference datasets should be incorporated after harmonization with the GACED30 definition; for example, samples from Lesiv et al. (2025) would require reinterpretation of woody crops and fallow or shifting-cultivation classes before use."

(2) The strength of the proposed dataset is the annually updated, inter-annual consistency. However, it is not clear how the proposed method ensure the consistency. Accordingly to the description in Line 227, the authors utilized year-specific CatBoost models to do binary classification. In this case, the CatBoost models trained in one year has no linkage with the next year or the

year before. It is not convincing that such annual mapped cropland be consistency across years.

> Thank you for raising this important distinction. You are correct that independently trained year-specific CatBoost models do not, by themselves, impose temporal coupling between adjacent annual maps. Our original wording did not adequately distinguish a harmonized annual production framework from explicitly enforced pixel-level interannual consistency.

The annual classifiers use the same harmonized SDC30 input framework, feature representation, expert-reference basis, sample-construction and screening protocol, and CatBoost configuration. These common elements reduce observational discontinuities and standardize the annual production process, thereby supporting cross-year comparability. However, they do not guarantee a temporally constrained classification trajectory, and the annual training subsets are not identical because augmented point-year labels can be excluded separately through year-specific CleanLab screening.

The only direct linkage between adjacent annual maps is the limited post-classification rotation rule in Sect. 3.3.1. This rule retains an uncertain one-year observation as Cropland only when the same pixel is confidently classified as Cropland in both the preceding and succeeding years. It therefore addresses a restricted class of isolated one-year gaps rather than smoothing or constraining the complete series. Similarly, the Mann–Kendall and Theil–Sen analyses operate on the completed annual maps and do not retrospectively constrain the annual classifications.

We have therefore replaced wording implying complete temporal consistency with the more accurate description of GACED30 as a harmonized annual baseline. Annual coverage and cross-year comparability remain important properties of the dataset, but they are no longer conflated with enforced pixel-level interannual consistency.

The manuscript has been revised as follows:

A. Title and Abstract, We replaced 'a consistent baseline' with 'a harmonized baseline' and stated that independent year-specific classifiers were trained under a common observation, feature, sample-construction, and model protocol.

B. Sects. 3.2 269–273 and 3.2.1, Lines 301–314 – We stated that the annual classifiers were independently trained, that direct temporal linkage was limited to Sect. 3.3.1, and that year-specific weak-label screening produced annual training subsets that shared a common framework but were not identical.

“To generate annual cropland probability maps under a common mapping framework, we first constructed spatially extensive and thematically harmonized annual training subsets through multi-stage sample alignment and year-specific weak-label screening (Sect. 3.2.1). We subsequently used year-specific CatBoost models to estimate pixel-wise cropland probabilities across the 2000–2024 archive (Sect. 3.2.2). The annual classifiers were trained independently, and direct temporal linkage between their outputs was introduced only through the limited post-classification rotation rule described in Sect. 3.3.1.”

“Finally, annual weak labels were assigned directly to the selected spatial samples for 2000–2024. For 2000–2019, the label from each GLAD Cropland epoch was assigned to every year covered by that epoch; because GLAD Cropland provides no subsequent epoch, the 2016–2019 labels were also used as the initial weak labels for 2020–2024. For each year, five-fold out-of-fold CatBoost probabilities were generated using StratifiedKFold(n_splits=5, shuffle=False) and supplied to CleanLab’s default find_label_issues procedure (Northcutt et al., 2021). FAST-Crop labels were treated as fixed reference labels and were excluded from possible pruning, whereas only the augmented point-

year labels were eligible for screening. No manually selected probability or confidence threshold was applied. An augmented point-year flagged by CleanLab was excluded from the training sample set for the corresponding annual classifier, and an augmented location was removed entirely when more than 25% of its 25 annual labels were flagged, corresponding to at least seven years. This procedure screens year-specific weak-label conflicts rather than imposing a transition rule between adjacent years. The annual training subsets therefore shared the same expert-reference basis, feature definition, sample-construction framework, and screening procedure, but were not identical because augmented point-year labels could be excluded separately for each annual classifier. After annual weak-label assignment and screening, the final augmented sample set contained 332,471 locations, comprising approximately 25,000 Cropland and 307,000 Non-cropland samples."

C. Sect. 3.2.2, Lines 326–334 – We explicitly stated that no parameter sharing, temporal regularization, or model-level coupling was imposed between adjacent annual models, and distinguished cross-year comparability from enforced consistency.

"The trained year-specific CatBoost models were applied to the corresponding annual SDC30 feature stacks to generate pixel-wise cropland probability maps for 2000–2024. Because these classifiers were fitted independently, no parameter sharing, temporal regularization, or explicit model-level coupling was imposed between adjacent years. Cross-year comparability is supported by the harmonized SDC30 input framework—which reduces observational discontinuities associated with missing observations and sensor transitions—and by the common feature representation, sample-construction protocol, and classification configuration. These measures standardize the annual mapping process but do not enforce complete pixel-level interannual consistency. Direct temporal evidence is introduced only through the limited post-classification

rotation rule described in Sect. 3.3.1. Production training and inference were performed on the high-performance iEarth remote-sensing cloud computing platform hosted by Pengcheng Laboratory."

D. Sect. 3.3.1, Lines 344–352 – We renamed and delimited the temporal rule, explaining that it targets only uncertain isolated one-year observations and does not impose general consistency across the annual series.

"Because the annual CatBoost models were trained independently, temporal evidence was introduced after classification through a limited rotation rule targeting uncertain observations. Specifically, we identified pixels within the probability interval ($P \in [0.3, 0.5]$), which would otherwise be classified as Non-cropland under the standard binary threshold.

For a pixel in year t, an uncertain observation within this interval was retained as Cropland under a temporary-fallow interpretation only when the same pixel was classified as active cropland (P>0.5) in both the preceding year t-1 and succeeding year t+1. This rule bridges isolated one-year gaps that are compatible with temporary fallow or weak phenological signals. The temporal context reduces, but does not eliminate, confusion with spectrally similar pasture, rangeland, or bare surfaces. The rule does not alter observations with probabilities below 0.3, smooth complete pixel trajectories, or impose general temporal consistency across the annual series."

E. Sect. 4.3.2, Lines **–** – We removed the circular interpretation of Cropping Frequency as independent evidence of temporal consistency and now present it as a descriptive indicator derived from the annual classifications themselves.

"Regions exhibiting a cropping frequency approaching 100% (represented by the darkest orange tones in Fig. 7) delineate areas with persistently mapped cropland status. These areas broadly coincide with established agricultural

regions, including the US Corn Belt, the North China Plain, and the Indo-Gangetic Plain. In these core production zones, the high frequency indicates that the corresponding pixels were classified as Cropland in most years under the common annual mapping protocol. Because cropping frequency is derived from the GACED30 classifications themselves, it should be interpreted as a descriptive temporal indicator rather than as independent validation of pixel-level interannual consistency. Conversely, extensive areas in rainfed agricultural regions, particularly the Sahel, Central Asia, and parts of Australia, exhibit intermediate cropping frequencies, typically between 40% and 70%. These values may reflect rotational systems, shifting cultivation, land-cover transitions, variable expression of fallow conditions, or residual annual classification variability. They should therefore be interpreted together with the Structural Evolution Indicators and complementary regional evidence."

F. Sects. 5.1, 5.3, and 6 – We discussed the basis and limits of cross-year comparability, acknowledged that complete pixel-level interannual consistency is not enforced, and consistently described GACED30 as a harmonized annual baseline in the Conclusion.

(3) Inter-comparison methodology in particular for area estimation based on different datasets are inconsistent. In Section 3.5.2, the authors correctly apply sample-based area estimation only to GACED30. While the area totals and trends for GLAD, ESA WorldCover, and GLC_FCS30D are derived from simple pixel counting, the GACED30's area is adjusted for omission/commission errors. A pixel-count comparison between a bias-corrected map (GACED30) and uncorrected maps is not a fair assessment.

> Thank you for identifying this methodological inconsistency. You are correct that the submitted comparison combined the sample-adjusted GACED30 area

with unadjusted mapped-area estimates for the peer products. Although the peer mapped areas were calculated using CRS-aware pixel areas rather than nominal pixel counts, they were not adjusted for their product-specific omission and commission errors. We agree that this did not provide a consistent basis for inter-product comparison.

We have therefore recalculated the area series of every map product shown in Fig. 6a using the same sample-based area-adjustment framework. For each product, we constructed a separate binary confusion matrix using the available layer or epoch closest to the FAST-Crop reference imagery time. For every available product year or epoch, its CRS-aware Cropland and Non-cropland area proportions were then combined with its own confusion matrix to derive the sample-adjusted cropland area, standard error, and 95% confidence interval. We also state the assumption that applying a fixed product-specific confusion matrix across a temporal series requires its error structure to remain reasonably stable over the corresponding coverage period.

The revised comparison therefore places all map products under a common adjustment framework while preserving differences in their native definitions, temporal coverage, and mapping behaviour. We have correspondingly regenerated Fig. 6 and rewritten Sect. 4.3.1 so that the trajectories are interpreted as product-specific estimates rather than a ranking based on adjusted versus unadjusted areas.

The manuscript has been revised as follows:

A. Sect. 3.5.4, Lines 501–510 – We replaced the GACED30-only adjustment with a common sample-based area-adjustment framework applied separately to every map product and stated the fixed-error-matrix assumption.

“To ensure a consistent inter-product comparison, we applied the same sample-based area-adjustment framework to all map products shown in Fig. 6a. The FAST-Crop reference set was generated using stratified random sampling

based on the underlying land-surface classes, together with a global equal-area hexagonal framework for spatial stratification. For each product, one product-specific binary confusion matrix was constructed using the available layer or epoch closest to the FAST-Crop reference imagery time. For every available year or epoch, the CRS-aware mapped Cropland and Non-cropland area proportions were combined with the corresponding product-specific confusion matrix to obtain sample-adjusted global cropland areas. Standard errors and 95% confidence intervals were calculated consistently following Olofsson et al. (2014) and Tyukavina et al. (2025). This procedure retains the product-specific omission and commission patterns while placing all area estimates relative to the same FAST-Crop reference definition. Applying a fixed confusion matrix across each product series assumes that its error structure remains reasonably stable over the corresponding temporal coverage."

B. Figure 6 – We regenerated all map-product trajectories using the consistently sample-adjusted estimates. The shaded uncertainty bands represent the corresponding 95% confidence intervals.

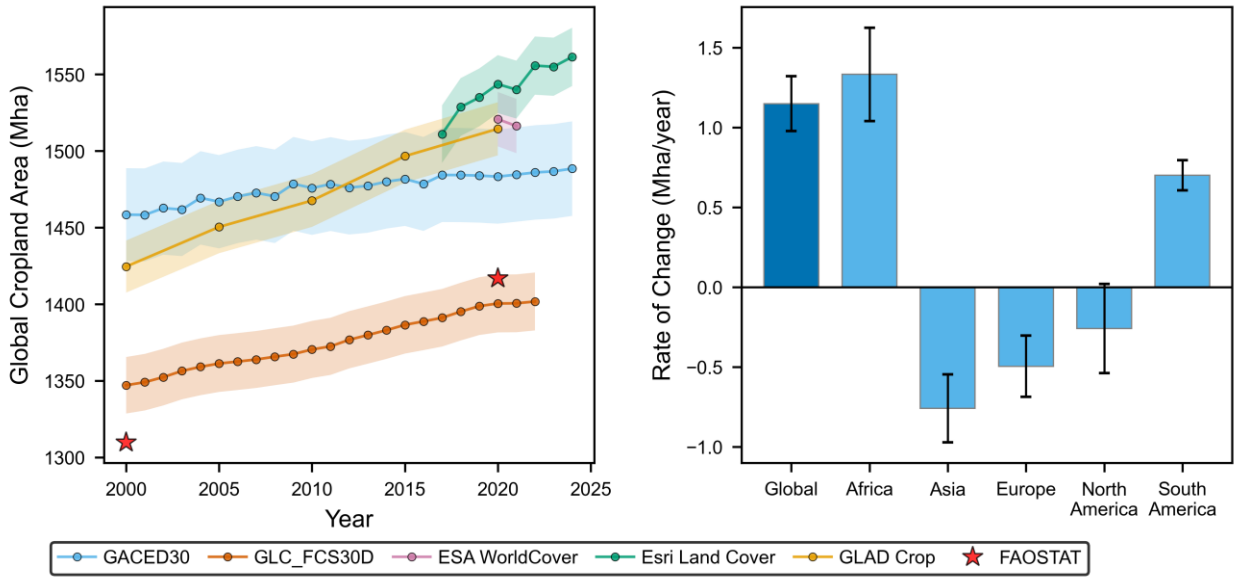

Figure 6. Inter-annual variation and trends in global cropland area from 2000 to 2024. (a) Global cropland-area trajectories from GACED30, GLC_FCS30D,

GLAD Cropland, Esri Land Cover, ESA WorldCover, and FAOSTAT. (b) Global and continent-level trends in cropland extent.

C. Sect. 4.3.1, Lines 702–719 – We replaced the comparison between adjusted and unadjusted area series with the consistently adjusted results and now interpret the remaining differences in relation to product-specific temporal coverage, definitions, and mapping behaviour.

"The sample-adjusted GACED30 series estimated global cropland area at 1458.5 Mha (95% CI: 1428.2–1488.8 Mha) in 2000, 1483.3 Mha in 2020, and 1488.5 Mha (1457.7–1519.3 Mha) in 2024 (Fig. 6a). This represents a net increase of approximately 30.0 Mha, or 2.1%, from 2000 to 2024, corresponding to an average increase of approximately 1.2 Mha year^{-1}.

The consistently adjusted peer-product series showed different area levels and endpoint changes over their respective periods. GLC_FCS30D increased from 1347.2 Mha in 2000 to 1401.8 Mha in 2022, whereas GLAD Cropland increased from 1424.6 Mha for the 2000–2003 epoch to 1514.5 Mha for the 2016–2019 epoch. Over its shorter temporal coverage, Esri Land Cover increased from 1511.1 Mha in 2017 to 1561.4 Mha in 2024. ESA WorldCover showed little change between its two available years, decreasing from 1520.8 Mha in 2020 to 1516.3 Mha in 2021. These results show that the adjustment does not force the products toward a common trajectory; substantial differences in both estimated area and temporal evolution remain. GACED30 provides a continuous 25-year annual record with a comparatively gradual long-term increase. Across products, the adjusted estimates form a continuum shaped by product-specific mapping behavior and native definitions, particularly their treatment of perennial woody crops, active fallow, temporary meadows and pastures, and mixed agricultural mosaics."

(4) The paper explicitly excludes "Temporary meadows and pastures" to avoid overestimation but includes "Active Fallow." In reality, differentiating a fallow field (bare soil) from a plowed temporary pasture or a degraded rangeland using 30-m Landsat time series is extremely challenging spectrally.

> Thank you for highlighting this observational challenge. We agree that temporary fallow, ploughed pasture, degraded rangeland, and natural bare or sparsely vegetated surfaces can exhibit similar spectral and phenological characteristics at 30-m resolution and cannot be perfectly separated in every landscape.

In GACED30, the inclusion of active fallow and exclusion of temporary meadows and pastures define the intended semantic scope of Cropland, but this definition does not guarantee complete pixel-level discrimination. The limited rotation rule retains an uncertain one-year observation as Cropland only when both adjacent years are confidently classified as Cropland. Topographic and temporal information can reduce confusion with pasture, rangeland, and natural bare surfaces, but cannot eliminate it.

We have therefore removed categorical wording such as 'successfully distinguishes,' 'correctly identifies,' and 'effectively filters out.' The revised manuscript describes the relevant features and temporal context as reducing ambiguity, treats the challenging-landscape comparisons as qualitative illustrations rather than direct validation, and explicitly acknowledges the remaining uncertainty in Sect. 5.3, particularly in semi-arid cropland-rangeland mosaics.

The manuscript has been revised as follows:

A. Sect. 2.1, Lines 108–127 – We clarified that active fallow is included as part of a cultivation rotation and that long-term phenological information is used to

reduce, rather than eliminate, confusion with abandoned or naturally barren land.

Please paste the complete cropland-definition passage beginning with:

“Unlike general land cover products that characterize the instantaneous physical state of the surface (e.g., bare soil or green vegetation) (Di Gregorio and Jansen, 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 as part of a cultivation rotation. Long-term phenological information was used to reduce confusion between temporary fallow and abandoned or naturally barren land.

Agricultural Structures: This includes greenhouses and other permanent production structures essential to modern agricultural systems.

While FAO includes "Temporary meadows and pastures," defined as land cultivated with forage for less than five years, these areas are often spectrally similar to natural grasslands in satellite imagery. Temporary meadows and pastures were excluded from the target cropland definition to reduce the inclusion of pasture and rangeland."

B. Sects. 2.2 Lines 137–141and 3.1, Lines 261–265 – We revised the descriptions of NASADEM and topographic features so that they provide contextual information and help reduce confusion without implying definitive separation.

"To provide additional topographic context, we integrated elevation, slope, and aspect derived from NASADEM (NASA JPL, 2020). These variables provide information on whether bare surfaces occur in terrain that is broadly compatible with cultivation and thereby help reduce confusion between agricultural bare land and natural barren areas in terrain less suitable for farming. The digital elevation model can be accessed via the NASA Earthdata portal at https://www.earthdata.nasa.gov/data/catalog/lpcloud-nasadem-hgt-001."

"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 the model distinguish flat arable land from natural vegetation on steep slopes, where extensive cultivation is generally less likely. Additionally, pixel-level Latitude and Longitude were included not merely as location variables, but to implicitly represent broad agro-climatic variation, allowing the model to adapt to geographical differences in crop calendars."

C. Sect. 3.3.1, Lines 344–352 – We replaced the categorical reclassification of Active Fallow with retention as Cropland under a temporary-fallow interpretation and explicitly stated the remaining ambiguity.

“Because the annual CatBoost models were trained independently, temporal evidence was introduced after classification through a limited rotation rule targeting uncertain observations. Specifically, we identified pixels within the probability interval ($P \in [0.3, 0.5]$), which would otherwise be classified as Non-cropland under the standard binary threshold.

For a pixel in year t, an uncertain observation within this interval was retained as Cropland under a temporary-fallow interpretation only when the same pixel was classified as active cropland (P>0.5) in both the preceding year t-1 and succeeding year t+1. This rule bridges isolated one-year gaps that are compatible with temporary fallow or weak phenological signals. The temporal context reduces, but does not eliminate, confusion with spectrally similar pasture, rangeland, or bare surfaces. The rule does not alter observations with probabilities below 0.3, smooth complete pixel trajectories, or impose general temporal consistency across the annual series.

D. Sect. 4.2.1, Lines 665–680 – We revised the Willamette Valley, Murray-Darling Basin, Puglia, and Gezira descriptions to report differences in mapped patterns without treating them as validation of fallow-pasture discrimination.

“To illustrate inter-product differences in challenging agricultural landscapes, we compared the 2020 GACED30 layer with temporally corresponding GLC_FCS30D and ESA WorldCover layers and the 2019 GLAD Cropland epoch (Fig. 4). The selected cases include cloud-prone, semi-arid, fragmented, and spectrally ambiguous agricultural environments and are intended to illustrate differences in mapped spatial patterns rather than provide case-specific accuracy validation.

In the Willamette Valley, USA (Fig. 4a), the products differ in their representation of managed grass-seed fields that may be spectrally similar to pasture and hay (Mueller-Warrant et al., 2011). Substantial differences are also visible in the drought-prone Murray–Darling Basin, Australia (Fig. 4d), where fallow cropland, pasture, and semi-arid vegetation can be difficult to distinguish (Xie et al., 2024). In Puglia, Italy (Fig. 4b), GACED30 maps broader cropland coverage across the mixed olive-grove and annual-crop landscape than GLAD Cropland, consistent with the inclusion of permanent woody crops in the GACED30 definition (Damianidis et al., 2021). In the Gezira Scheme, Sudan (Fig. 4c), the products differ in their treatment of temporarily bare fields within the irrigated agricultural landscape.

Differences are also apparent in the fragmented paddy and aquaculture landscape of the Ganges Delta, Bangladesh (Fig. 4e), and in the persistently cloud-prone Chocó Department, Colombia (Fig. 4f), where GACED30 produces a comparatively continuous mapped cropland pattern. These examples demonstrate how product definitions, observation strategies, and temporal representations influence mapped cropland patterns; without independent case-specific reference data, they should not be interpreted as quantitative evidence that one product is universally more accurate."

E. Sect. 5.3, Lines 859–867 – We expanded the limitations to state that temporary fallow, ploughed pasture, degraded rangeland, and natural bare or sparsely vegetated surfaces may remain spectrally and phenologically similar at 30 m.

"The principal spatial and semantic limitations of GACED30 arise from its binary class structure and 30-m spatial resolution. Active cultivation and rotational fallow are represented within a single Cropland class, so the dataset does not distinguish their annual management status. Fields smaller than or comparable to a Landsat pixel may be omitted or spatially smoothed, particularly in

fragmented smallholder landscapes. Temporary fallow, ploughed pasture, degraded rangeland, and natural bare or sparsely vegetated surfaces can also exhibit similar spectral and phenological characteristics. The AEZ assessment illustrates this geographical heterogeneity: in the cool, humid tropical stratum, the producer's accuracy of 0.498 indicates substantial cropland omission despite a high overall accuracy. The boreal sub-humid and humid reference subsets contained no positive Cropland cases and therefore did not support class-specific evaluation. Local applications in these environments require additional regional validation."

(5) Validation needs substantially enhanced. On one side, the validation shall targets not only for one year but also the 25 years. On the other side, validation over the inter-annual changes (expansion or reduction areas) shall be amended. While comparison with FAO statistical data provides quantitative accuracy metrics, it does not mean GACED30 outperformed other cropland layers or products are the metrics are a mixture of the classification accuracy and the discrepancy of definition of the classification schemes.

> Thank you for identifying the insufficient temporal scope of the submitted validation, the lack of a targeted assessment of cropland transitions, and the overinterpretation of the FAOSTAT comparison. We agree that the stable-site FAST-Crop assessment primarily evaluates static thematic agreement and does not validate all 25 annual layers or crop gain and crop loss.

We have therefore added independent multi-temporal transition assessments using the GLAD Cropland and LUCAS reference samples. The GLAD assessment evaluates four transitions between adjacent four-year epochs spanning 2000-2019, while the LUCAS assessment uses five successive survey intervals spanning 2006-2022. GACED30 and GLC_FCS30D were evaluated using identical reference observations, temporal aggregation or pairing procedures, and harmonized binary class rules.

GACED30 achieved higher recall and F1 for crop gain and crop loss in both assessments, although its GLAD-based precision was slightly lower and the absolute transition scores remained modest. We therefore interpret these results as common-sample evidence of greater sensitivity to the evaluated transitions rather than definition-independent proof of superiority.

We also explicitly acknowledge that the available references do not provide complete independent validation of every annual layer from 2000 to 2024. GLAD represents four-year epochs, and LUCAS observations are separated by three or four years. These assessments therefore evaluate agreement in the transition occurring over an interval, rather than independently identifying its exact annual timing or all within-interval reversals.

Finally, we have reframed the FAOSTAT comparison as an assessment of national-scale area consistency and long-term directional agreement. It does not measure pixel-level classification or transition accuracy and cannot by itself establish that GACED30 outperforms other cropland products, because statistical agreement remains influenced by differences in cropland definition, temporal coverage, and spatial support.

The manuscript has been revised as follows:

A. Abstract and Sect. 1 We expanded the evaluation summary to distinguish static thematic assessment, multi-temporal transition assessment, matched regional evaluation, sample-adjusted area comparison, and national statistical consistency.

B. Sects. 2.3.2–2.3.3, Lines 171–198 – We clarified the stable-site emphasis of FAST-Crop and added the independent GLAD Cropland and LUCAS multi-temporal reference sources, including their temporal coverage and data access.

“To provide an objective static thematic assessment of GACED30, we employed a validation set isolated entirely from model calibration. This subset

was extracted from the temporally stable records of the FAST validation set (Li et al., 2017). The FAST project generated its validation data using a sampling protocol distinct from that used for the training data: whereas training sites were manually selected for spectral representativeness, validation samples were allocated through a probabilistic global equal-area random sampling design. This probability-based framework provided spatially distributed observations of global land-cover heterogeneity. Unlike the augmented training samples, these validation samples were not subject to GLAD-based stratification or automated agreement filtering, thereby preventing GLAD-conditioned candidate selection from being propagated into this assessment. The final validation set comprised 29,226 sample locations, including 3,268 confirmed stable cropland samples. Because this subset deliberately emphasizes stable land-cover states, it was used to evaluate thematic agreement near the reference observation date but was not interpreted as direct validation of cropland gain, crop loss, or transition timing."

"To complement the static FAST-Crop assessment, we used two external multi-temporal reference sources whose reference labels were not used in GACED30 training or spatiotemporal post-processing. The first was the global probability-based reference sample associated with GLAD Cropland, comprising 3,500 sample locations with cropland interpretations for five four-year epochs: 2000–2003, 2004–2007, 2008–2011, 2012–2015, and 2016–2019 (Potapov et al., 2022). These observations supported the assessment of crop gain and crop loss between four pairs of adjacent epochs.

The second reference source was the European Union Land Use/Cover Area frame Survey (LUCAS), which provides harmonized in situ land-cover observations for the 2006, 2009, 2012, 2015, and 2018 survey campaigns (d'Andrimont et al., 2020), supplemented with observations from the 2022 campaign (d'Andrimont et al., 2024). Only locations with valid observations at both endpoints of a survey interval were retained.

GLAD Cropland provides global transition evidence across four adjacent four-year intervals, whereas LUCAS supplies denser repeated observations over five European survey intervals. Together, these datasets enable a targeted multi-temporal assessment of crop gain and crop loss, although their temporal spacing does not independently identify the exact year in which a transition occurred.

The GLAD Cropland reference samples are available at https://glad.umd.edu/dataset/croplands. The LUCAS primary data are available from https://ec.europa.eu/eurostat/web/lucas/database/primary-data."

C. Sect. 3.5, Lines 401–407 – We reorganized the validation framework into five complementary components and explained the distinct evidence provided by each component.

"To assess the quality and reliability of GACED30, we adopted five complementary approaches (Fig. 2e): (1) static thematic accuracy assessment using the independent FAST-Crop validation samples; (2) multi-temporal assessment of crop gain and crop loss using GLAD Cropland and LUCAS reference samples; (3) matched regional evaluation in China together with agro-ecological-zone-stratified accuracy assessment; (4) global and regional area estimation and inter-comparison with existing global and regional cropland products; and (5) comparison with national agricultural statistics from FAOSTAT. These components respectively evaluate static thematic agreement, transition detection across multiple intervals, regional and environmental variations in performance, sample-adjusted area trajectories, and aggregated national-scale consistency."

D. Sect. 3.5.2, Lines 445–470 – We added the temporal aggregation, adjacent-epoch comparison, repeated-observation pairing, class-harmonization, pooling, and transition-metric procedures for GLAD Cropland and LUCAS.

“We evaluated two directional transitions separately: Cropland to Non-cropland, hereafter termed crop loss, and Non-cropland to Cropland, hereafter termed crop gain. For each directional assessment, the target transition was treated as the positive class, whereas stable pairs and the opposite transition were treated as negative cases. These sample-level transitions provide a targeted assessment of change detection but are not identical to the 1-km structural expansion and reduction classes, which are derived from trends fitted to the complete 25-year annual record.

For the GLAD Cropland assessment, the annual binary layers of GACED30 and GLC_FCS30D were aggregated within each corresponding four-year epoch using the pixel-wise median. Crop gain and crop loss were then determined for each of the four adjacent-epoch intervals: 2000–2003 to 2004–2007, 2004–2007 to 2008–2011, 2008–2011 to 2012–2015, and 2012–2015 to 2016–2019. The mapped transitions were evaluated against the corresponding GLAD reference transitions at the same sample locations. Each valid location–interval pair was treated as one observation, and the four intervals were pooled for the overall assessment. The same temporal aggregation, spatial extraction, and class-harmonization procedures were applied to GACED30 and GLC_FCS30D.

For LUCAS, valid observations at the same location in successive available surveys were paired. The corresponding annual GACED30 and GLC_FCS30D classes were extracted for the two survey years, and the reference and mapped transitions were determined from their respective endpoint classes. When a location occurred in more than one survey interval, each location–interval pair was treated as one unweighted observation. Results from the 2006–2009,

2009–2012, 2012–2015, 2015–2018, and 2018–2022 intervals were pooled for the overall assessment.

Class harmonization was performed before determining the transitions. GACED30 class 10 was treated as Cropland. For LUCAS, land-cover codes B, B11–B19, B21–B23, B31–B37, B41–B45, B51–B54, B71–B77, B81–B84, BX1, and BX2 were mapped to Cropland. Temporary grassland (B55) was mapped to Grassland and was therefore excluded. For GLC_FCS30D, classes 10, 11, 12, and 20 were mapped to Cropland, whereas class 130 was treated as Grassland. Consequently, permanent crops were included, temporary grasslands were excluded, and fallow land was not automatically assigned to Cropland but followed its original observed or product land-cover class.

Because stable observations predominated in both reference sources, overall accuracy was not used as the principal transition metric, as it would be dominated by unchanged pairs. The resulting metrics were interpreted within the spatial, temporal, and semantic support of each reference dataset."

E. Sect. 4.1.1 and Table 2, Lines 544–558 – We renamed the analysis 'Static thematic accuracy assessment' and clarified that the metrics measure agreement with the adopted FAST-Crop/GACED30 definition rather than temporal validation or universal product superiority.

Table 2: Static thematic accuracy of GACED30 and peer products against the independent FAST-Crop stable-site validation set

Product	F1	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	0.570
GLC_FCS10	0.657	0.917	0.633	0.683	0.610
ESA WorldCover	0.722	0.941	0.790	0.664	0.689

Esri Land Cover	0.689	0.931	0.727	0.655	0.651

“To assess static thematic agreement, we evaluated GACED30 using the independent FAST-Crop stable-site validation set, which was excluded from model training. We then compared GACED30 with six global 10 and 30 m land-cover or cropland products: GLAD Cropland, GCEP30, GLC_FCS30D, GLC_FCS10, ESA WorldCover, and Esri Land Cover.

As shown in Table 2, GACED30 achieved the highest OA (0.965), F1 score (0.844), UA (0.883), and PA (0.808) among the evaluated products. Its F1 score exceeded those of GLAD Cropland and ESA WorldCover by 0.100 and 0.122, respectively. These results demonstrate strong static thematic agreement between GACED30 and the independent FAST-Crop validation samples under the adopted Cropland/Non-cropland definition.

Part of the mismatch with peer products may reflect differences in cropland definitions rather than mapping errors alone. Specifically, GACED30 includes orchards and active fallow but excludes temporary meadows and pastures, whereas peer products differ in their treatment of these categories. Temporal mismatches between the validation samples and product layers may also contribute to the observed differences. The metrics in Table 2 should therefore be interpreted as the products' relative agreement with the adopted FAST-Crop/GACED30 definition rather than as evidence of universal superiority across all cropland categories. Accordingly, Table 2 supports the static thematic reliability of GACED30 under the adopted Cropland/Non-cropland reference definition but does not directly assess its ability to detect inter-period cropland changes.”

F. Sect. 4.1.2 and Table 3, Lines 577–598 – We added the multi-temporal crop-gain and crop-loss results, including the higher recall and F1 of GACED30, its slightly lower GLAD-based precision, and the modest absolute transition scores.

Table 3: Pooled transition detection by GACED30 and GLC_FCS30D using the GLAD Cropland and LUCAS multi-temporal reference samples

Reference set	Target transition	Product	Precision	Recall	F1
GLAD Cropland	Cropland → Non-cropland	GACED30	0.180	0.177	0.178
	Cropland → Non-cropland	GLC_FCS30D	0.186	0.129	0.152
	Non-cropland → Cropland	GACED30	0.227	0.199	0.212
	Non-cropland → Cropland	GLC_FCS30D	0.232	0.171	0.197
LUCAS	Cropland → Non-cropland	GACED30	0.099	0.072	0.083
	Cropland → Non-cropland	GLC_FCS30D	0.062	0.016	0.026
	Non-cropland → Cropland	GACED30	0.113	0.058	0.076
	Non-cropland → Cropland	GLC_FCS30D	0.057	0.014	0.023

"The multi-temporal assessment (Table 3) provides a targeted evaluation of crop gain and crop loss that is not available from the stable-site FAST-Crop samples. Across the four pooled adjacent GLAD Cropland epoch intervals, GACED30 achieved higher recall and F1 score than GLC_FCS30D for both transition directions. For crop loss, its recall and F1 score were 0.177 and 0.178, compared with 0.129 and 0.152 for GLC_FCS30D. For crop gain, the corresponding values were 0.199 and 0.212 for GACED30 and 0.171 and 0.197 for GLC_FCS30D. GACED30 nevertheless had slightly lower precision: 0.180 versus 0.186 for crop loss and 0.227 versus 0.232 for crop gain.

The pooled LUCAS comparison showed larger differences between the products. For crop loss, GACED30 achieved precision, recall, and F1 score values of 0.099, 0.072, and 0.083, respectively, compared with 0.062, 0.016, and 0.026 for GLC_FCS30D. For crop gain, GACED30 obtained values of 0.113, 0.058, and 0.076, whereas GLC_FCS30D obtained 0.057, 0.014, and 0.023.

GACED30 therefore achieved higher transition recall and F1 score in both reference assessments, indicating greater sensitivity to the evaluated reference transitions. However, the absolute metrics remained modest, and its GLAD-based precision was slightly lower than that of GLC_FCS30D. Transition detection requires correct classification at both temporal endpoints, so an error at either endpoint can generate an incorrect change label. The low prevalence of transitions, temporal aggregation, spatial-support differences, and residual semantic differences further affect these metrics. The results provide a controlled common-sample comparison of transition detection across multiple intervals but do not establish definition-independent superiority or validate the exact year of annual changes."

G. Sect. 4.1.4 Lines 624–637 – We revised the FAOSTAT interpretation to distinguish national-scale area consistency and long-term directional agreement from pixel-level thematic or transition accuracy.

"To evaluate national-scale area consistency, we aggregated the GACED30 classifications into national totals and compared them with definition-reconciled FAOSTAT statistics. Across 132 countries, GACED30 achieved an R2 of 0.95 and an NRMSE of 3.7% (Fig. 3a). These metrics indicate close agreement in country-level cropland area after definition reconciliation but do not measure pixel-level thematic accuracy or transition detection.

For broader context, these values can be compared with those reported for other global cropland products in Table D1 of Tubiello et al. (2023a). Because those values were obtained using different product years, cropland definitions, and comparison protocols, they provide contextual information on the magnitude of national-area agreement rather than a controlled inter-product ranking. The sample-based assessments in Sects. 4.1.1–4.1.3 provide the more direct comparisons of static thematic, multi-temporal transition, and regional performance.

Beyond national area levels, we compared the long-term net change derived from GACED30 with FAOSTAT for 42 countries reporting substantial cropland dynamics (Fig. 3b). The directional match rate was 73.9%, and the area-weighted match rate was 83.1%. These results support agreement in the aggregated direction of long-term cropland change, particularly for major agricultural countries. Because the comparison is based on national totals, it does not validate the spatial location of changes, their exact annual timing, or their underlying causes."

H. Sect. 4.3.3, Lines 794–803 – We clarified that GLAD adjacent-epoch transitions and the GACED30 full-record Structural Evolution Indicator provide complementary representations of change and should not be treated as directly equivalent validation targets.

"Comparison with the GLAD Cropland change maps shows both broad agreement and systematic differences between the two analytical approaches. Both products identify major expansion frontiers in South America and Africa, whereas larger discrepancies occur in semi-arid and transitional regions such as Central Asia and Australia. GLAD represents crop gain and crop loss between adjacent four-year epochs, while the GACED30 Structural Evolution Indicator identifies persistent monotonic trends across the complete annual record under specified slope and significance criteria. Consequently, a change detected within an individual GLAD epoch interval may not produce a statistically supported monotonic trend over the complete 2000–2024 GACED30 record. This difference does not by itself establish that either classification is erroneous. The two products instead provide complementary representations of cropland dynamics: GLAD characterizes changes between consecutive multi-year epochs, whereas GACED30 provides a more conservative delineation of persistent structural trends across the full annual series."

I. Sects. 5.1 Lines 820–826 and 5.3, Lines 868–875 – We discussed the distinct evidence supplied by the complementary assessments and acknowledged that the available references cannot validate every annual layer, exact transition year, or all within-interval reversals.

“The complementary evaluations describe different aspects of product performance. The FAST-Crop assessment indicates strong static thematic agreement, the matched China assessment shows overall accuracy comparable with CACD, and the AEZ-stratified results identify geographically heterogeneous performance. The GLAD Cropland and LUCAS assessments provide direct evidence on crop-gain and crop-loss detection across multiple intervals: GACED30 achieved higher recall and F1 than GLC_FCS30D on the common samples, although the modest absolute scores demonstrate the continuing difficulty of transition mapping. Together, these results support the use of GACED30 for global and regional structural monitoring but do not constitute validation of the exact year or cause of every local change. performance.”

“Temporal comparability is supported by the common observation, feature, sample-construction, and model protocols, but complete pixel-level interannual consistency is not enforced. The independently trained annual classifiers can remain sensitive to annual spectral conditions, particularly near the classification threshold or during abrupt land-cover changes. Differences between the earlier Landsat–MODIS observation environment and the more recent sensor era may also influence temporal comparability and are not eliminated solely by trend analysis. Moreover, the available GLAD Cropland and LUCAS references assess crop gain and crop loss across multi-year intervals rather than every annual layer. They cannot identify the exact transition year, detect all within-interval reversals, or distinguish permanent

cropland reduction from temporary non-cropland states. More independent annual and transition-specific reference observations are therefore needed."

J. Sect. 6, Lines 894–898 – We added a concise, qualified conclusion concerning static and multi-temporal performance and the remaining difficulty of local annual transition mapping.

"Independent stable-site validation yielded an overall accuracy of 0.965 and an F1 score of 0.844, while the matched assessment in China showed that GACED30 and CACD both achieved an overall accuracy of 0.94 under the same reference-sample protocol. Multitemporal assessments yielded higher recall and F1 scores than GLC_FCS30D for both crop gain and crop loss; however, the modest absolute transition scores indicate that local annual changes should be interpreted cautiously."

Minor Comments

(1) In Table 1 - Temporal attributes for "ESA WorldCover" is "Annual 2000-2001". This appears to be a typo.

> Thank you for identifying this typographical error. We have corrected the temporal coverage of ESA WorldCover from 2000-2001 to 2020-2021 in Table 1.

(2) Line 315 (Section 3.4). The slope threshold of > 0.2 ha/year is defined. In a 1 km grid, 0.2 ha is roughly 2.2 Landsat pixels per year. Given that the product is 30-m resolution, is this threshold reasonable to define whether such changes are not "minor fluctuation"?

> Thank you for this careful observation. You are correct that 0.2 ha is numerically equivalent to approximately 2.2 nominal 30-m pixel areas and that

the submitted wording could be interpreted as a consecutive-year pixel-count threshold distinguishing genuine change from minor fluctuation. We have removed that interpretation and clarified how the threshold is applied.

The threshold is applied to the Theil–Sen slope estimated from the complete 25-observation cropland-area trajectory within each 1 × 1 km grid cell, rather than to the number of 30-m pixels changing between two consecutive years. A slope of 0.2 ha yr^{-1} corresponds to a linear-equivalent change of 4.8 ha, or approximately 4.8% of the grid area, over the 24 elapsed annual intervals. We now describe it as an operational grid-scale effect-size threshold rather than a sensor-derived detection limit, an empirical estimate of classification noise, or a universally optimal boundary.

Because neither the Mann–Kendall test nor the Theil–Sen estimator provides a theoretically unique value of the minimum slope threshold, we also evaluated thresholds from 0.10 to 0.50 ha yr^{-1}. The exact mapped extent changes with the threshold, as expected, but values close to 0.2 retain high spatial agreement and similar continental composition. We therefore retain 0.2 ha yr^{-1} for the principal analysis while documenting its operational interpretation, limitations, and sensitivity.

The manuscript has been revised as follows:

A. Sect. 3.4, Lines 383–388 – We explained that the threshold is applied to the complete 25-year grid-level trajectory, quantified its study-period equivalent, identified it as operational rather than theoretically optimal, and introduced the sensitivity analysis.

“Neither the MK test nor the Theil–Sen estimator provides a theoretically unique or universally optimal value of τ. The appropriate minimum effect size depends on the spatial aggregation unit, study duration, and analytical objective. For the principal analysis in this study, we adopted τ=0.2 ha yr^{-1} as an operational grid-scale threshold. Because a 1 × 1 km grid cell covers approximately 100 ha, this

value corresponds to a linear trend equivalent to 0.2% of the grid area per year and to 4.8 ha, or 4.8% of the grid area, across the 24 elapsed annual intervals between 2000 and 2024. It therefore provides an interpretable minimum magnitude for identifying landscape-scale structural change over the study period.

Although 0.2 ha is numerically equivalent to approximately 2.2 nominal 30 m pixel areas, τ is not applied as a single-year pixel-count detection limit. Instead, β is estimated from the complete 25-year cropland-area trajectory within each grid cell. We do not consider τ=0.2 ha yr^{-1} to be universally optimal, and other thresholds may be appropriate for analyses with different objectives or tolerances for lower-magnitude change. We therefore evaluated the sensitivity of both the general and stringent structural-evolution classifications by varying (τ) from 0.10 to 0.50 ha yr-1 at intervals of 0.05 ha yr-1 (Figs. S1 and S2).”

B. Sect. 4.3.3, Lines 775–786 – We added the sensitivity results for τ = 0.10–0.50 ha yr^{-1}, including changes in global class extent, continental composition, and spatial agreement with the τ = 0.2 masks.

“Because τ is an operational rather than theoretically optimal threshold, we examined whether the principal geographical interpretation depended strongly on its selected value. Increasing τ from 0.10 to 0.50 ha yr−1 progressively reassigned lower-magnitude expansion and reduction grid cells to stable, as expected, and therefore changed the absolute mapped extent of structural change (Fig. S1a). However, the continental composition of the general expansion and reduction areas changed only modestly across the evaluated thresholds, with no abrupt redistribution around τ=0.2 ha yr−1 (Fig. S1b–c). Spatial agreement with the τ=0.2 reference masks was also high for nearby thresholds under both definitions: at τ=0.15 and 0.25, the Jaccard indices were approximately 0.88–0.91 for the general masks and 0.91–0.93 for the stringent

masks (Fig. S2). Thus, the precise extent of lower-magnitude change remains threshold-dependent, but the broad continental distribution and principal spatial patterns are retained across reasonable values surrounding τ=0.2. Together with the regional patterns in Fig. 8, this supports the robustness of the central interpretation that agricultural expansion is disproportionately concentrated in the Global South, whereas widespread stability and localized contraction characterize much of the Global North. Nevertheless, the sensitivity results do not imply that τ=0.2 is a uniquely optimal threshold."

C. Figs. S1 and S2 – We added supplementary analyses of trend-class composition, continental distribution, and Jaccard overlap across the evaluated thresholds.

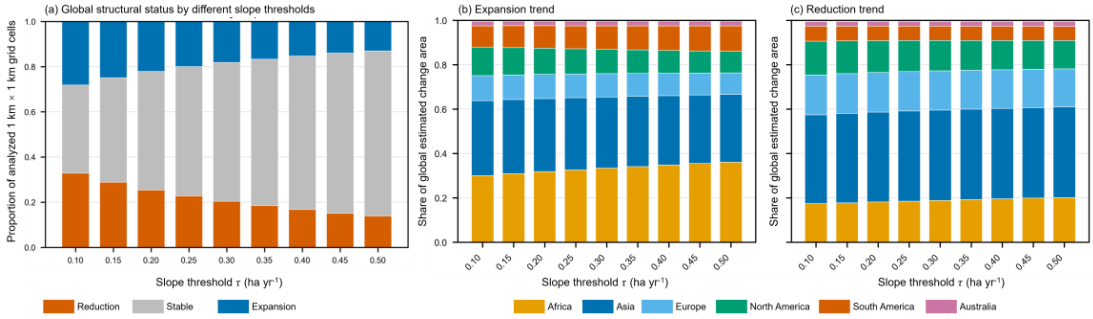

Figure S1. Sensitivity of global structural evolution trend composition and the continental distribution of cropland change to the minimum slope threshold. (a) Proportions of analysed 1 × 1 km grid cells assigned to general expansion ($\beta>\tau$), general reduction ($\beta<-\tau$), and general stable ($|\beta| \leq \tau$) as the minimum slope threshold τ varies from 0.10 to 0.50 ha yr−1. This panel applies the general structural trend definition independently of MK significance. (b–c) Continental shares of the total global area classified as expansion and reduction, respectively, under the corresponding thresholds. For each threshold, the continental shares within each change class sum to 1. Panels (b–c) describe the continental composition of the globally detected change area.

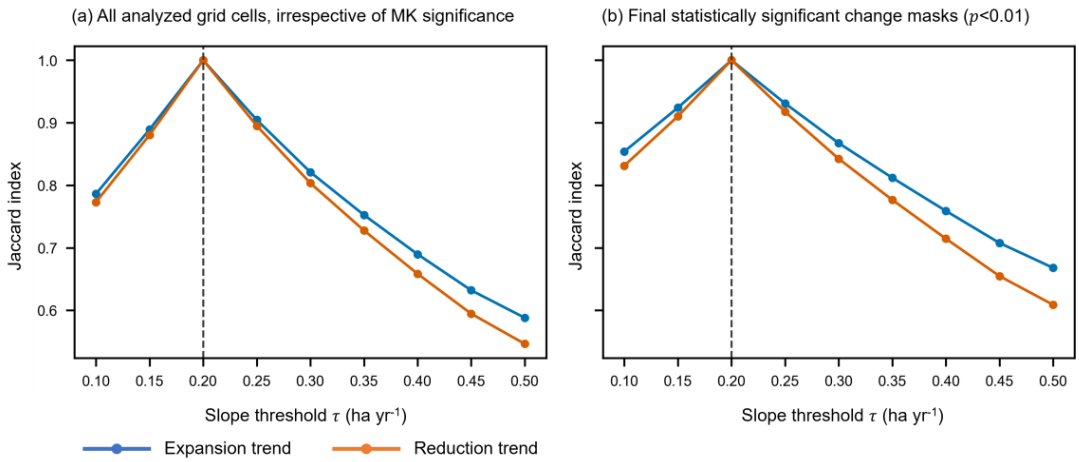

Figure S2. Sensitivity of the spatial expansion and reduction masks to the minimum slope threshold. The conventional Jaccard index, $J = |M_\tau \cap M_{0.2}| / |M_\tau \cup M_{0.2}|$, compares the mask M obtained at each minimum slope threshold τ with the corresponding reference mask at τ=0.2. Results are shown for (a) all analysed 1 × 1 km grid cells classified according to the sign and magnitude of the Theil–Sen slope, irrespective of Mann–Kendall significance, and (b) the final statistically significant change masks, for which expansion requires p<0.01 and β>τ, and reduction requires p<0.01 and β<-τ. A Jaccard index of 1 indicates identical masks. Panel (b) corresponds to the results of the statistically significant change masks.

D. Sect. 5.1, Lines 813–819 – We clarified that the selected threshold controls the interpreted minimum effect size but is neither uniquely optimal nor an empirical measurement of classification noise.

"The complete annual record supports two complementary descriptions of cropland dynamics. Cropping Frequency summarizes the proportion of years in which a pixel was mapped as Cropland, whereas the Structural Evolution Indicators use annual grid-level cropland area, the Mann–Kendall test, and Theil–Sen slopes to identify persistent directional patterns. Analysis of the full trajectory reduces sensitivity to the selection of individual endpoint years, but the exact spatial extent of expansion and reduction remains dependent on the minimum slope threshold. The sensitivity analysis indicates that the broad continental distribution is retained across thresholds near τ=0.2 ha yr−1, without

implying that this value is uniquely optimal or that the detected trends establish their underlying causes."

(3) Figure 5 title states "GACED30 (2000-2022)" but the paper's title and dataset span 2000-2024. Please align the figure period with the full dataset period or explain why 2023-2024 is omitted from this specific visualization. Also, why the authors presented different conversion periods on the map of Brazil, China, and Kaz.

> Thank you for identifying this ambiguity. We have corrected the Fig. 5 caption to state explicitly that the GACED30 dataset covers 2000–2024. Figure 5 is not intended to display a common temporal subset or imply that data for 2023–2024 are unavailable. Instead, each panel uses a region-specific interval selected to encompass its principal visually evident cropland-status change episode: 2010–2022 for expansion in Matopiba, Brazil; 2000–2015 for reduction in the Yangtze River Delta, China; and 2000–2010 for reduction in the Aral Sea basin, Kazakhstan. We have also clarified that the colours indicate the first mapped year within the displayed interval and that the examples do not independently verify the conversion date, cause, or comparability of change magnitude among regions.

(4) As presented in Figure 7 a, several non-agricultural areas or minor agriculture producing regions presented statistical significance of changing trends. Is this the consequences of classification bias or the actual phenomenon, for example, the significant changing trends in Sahara desert.

> Thank you for identifying this problem. After re-examining the submitted global trend panel, we found that the apparent signals in non-agricultural regions, including the Sahara Desert, were caused by an incorrect layer used during figure preparation. We replaced it with the correct output layer and

regenerated the global panel. The revised figure, now numbered Fig. 8a, shows the expected spatial distribution of cropland trends and no longer contains the anomalous Sahara signals. Because this was a figure-layer error rather than a change to the trend-analysis method, no further methodological modification was required.

(5) In Data Availability section, I expect the authors focusing on the products generated by this paper or the key inputs data produced by the team, while removing the information of other source of dataset. Or, you could document the way of data access for other land cover products in the Data part.

> Thank you for this helpful suggestion. We agree that the Data Availability section should focus on the product generated in this study and the core input and reference datasets produced or maintained by our team. We have therefore retained only the access information for GACED30, SDC30, and FAST-Crop in this section and removed the previous itemized descriptions of third-party datasets. Access information for external products and validation sources is now provided in the relevant Dataset subsections.

The manuscript has been revised as follows:

A. Data availability, Lines 905–909 – We streamlined this section to include GACED30 and the SDC30 and FAST-Crop datasets used as the core input and reference sources.

“The Global 30 m Annual Cropland Extent Dynamics (GACED30) dataset generated in this study is openly available from Zenodo at https://doi.org/10.5281/zenodo.18199675 (Chen et al., 2026). Each annual raster layer follows the FROM-GLC coding convention, in which a value of 10 denotes cropland and a value of 0 denotes non-cropland. The SDC30 data cube and the FAST-Crop sample set, which were used as the core input and

reference datasets for GACED30 generation and validation, are available at https://data-starcloud.pcl.ac.cn/.”