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.