



Supplement of

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

Yuanhong Liao et al.

Correspondence to: Yuqi Bai (yuqibai@gmail.com), Jie Wang (wangj10@pcl.ac.cn), and Peng Gong (penggong@hku.hk)

The copyright of individual parts of the supplement might differ from the article licence.

S1. Supplementary methods and experiments for training-sample construction and classifier

S1.1 Augmented-sample construction and annual weak-label screening

The augmented sample set was constructed through four sequential stages.

Stage 1: FAST-Crop-based agreement screening. The initial candidate locations were generated through global equal-area sampling stratified by GLAD Cropland, with additional sampling in under-represented ecoregions. A preliminary CatBoost classifier trained exclusively with the expert-annotated FAST-Crop training samples was applied to these candidates. A candidate was retained only when its predicted Cropland/Non-cropland class agreed with the corresponding GLAD-stratified label. FAST-Crop therefore provided the semantic reference for agreement screening, whereas GLAD Cropland provided the initial spatial and thematic stratification prior.

Stage 2: spatial, spectral, and category filtering. The global land surface was divided into $5^\circ \times 5^\circ$ geographic grids to account for regional differences in surface-reflectance characteristics and sample composition. Within each grid, the six SDC30 surface-reflectance bands—Blue, Green, Red, NIR, SWIR1, and SWIR2—were transformed as $z = \log(1 + SR)$. This transformation was motivated by the statistical property that surface reflectance in homogeneous land-cover regions follows a Gamma distribution, while its logarithm approximately follows a Gaussian distribution under high-homogeneity conditions (Liao et al., 2026a). FAST-Crop cropland samples were used to estimate the corresponding multivariate Gaussian reference distribution. For an augmented cropland candidate with transformed reflectance vector ($\mathbf{z}$), the squared Mahalanobis distance was calculated as

$$D^2 = (\mathbf{z} - \boldsymbol{\mu}_{cg})^T \boldsymbol{\Sigma}_{cg}^{-1} (\mathbf{z} - \boldsymbol{\mu}_{cg})$$

where $\boldsymbol{\mu}_{cg}$ and $\boldsymbol{\Sigma}_{cg}$ denote the mean vector and covariance matrix of the FAST-Crop cropland samples in the corresponding geographic grid or its prescribed fallback distribution. A candidate passed the hard spectral prefilter only when $D^2 \leq \chi_{0.999,6}^2 = 22.457744$. A candidate was rejected when no reliable exact-grid, parent-grid, or global Gaussian reference distribution was available, when $D^2 > 22.457744$, or when its log-likelihood was non-finite. Candidates passing the hard prefilter were sampled according to their Gaussian likelihood using a likelihood temperature of 1. Finally, category balancing was conducted separately within each grid by sampling the filtered augmented candidates according to the Cropland/Non-cropland proportions represented by FAST-Crop in that grid.

Stage 3: performance-guided sample optimization. The samples retained after Stage 2 were further optimized using the sample-size–performance scaling relationship established by Liao et al. (2026b). Model performance as a function of training-sample size was represented using the log-logistic formulation

$$P(N) = P_{\infty} + \frac{P_0 - P_{\infty}}{1 + a(N-1)^k}$$

where N is the training-sample size, $P(N)$ is the estimated classification performance at N , P_0 is the initial performance, P_{∞} is the estimated asymptotic performance, k controls the rate at which performance approaches its asymptotic limit, and a controls the sample-size scale of the transition from the cold-start to saturation regime. At each iteration, ten alternative category-balanced batches of 10 000 samples were randomly drawn from the remaining candidate pool using the category-balancing rule described in Stage 2. Each candidate batch was provisionally combined with the FAST-Crop training samples and the augmented samples retained during previous iterations, and the sample-size–performance relationship was fitted to estimate P_{∞} . The batch producing the highest estimated P_{∞} was retained and removed from the candidate pool. The remaining candidate batches were returned to the pool.

Stage 4: annual weak-label extension and screening. Each retained augmented location was assigned annual weak labels for 2000–2024 using the class of the temporally nearest GLAD Cropland epoch. For each year, out-of-fold CatBoost probabilities were generated through five-fold stratified cross-validation using StratifiedKFold($n_splits=5$, $shuffle=False$). These probabilities were supplied to CleanLab’s default `find_label_issues` function, with no manually specified confidence threshold. FAST-Crop labels were treated as fixed and were not eligible for pruning. An augmented point-year flagged as a likely label issue was excluded from the training sample set of the corresponding annual classifier. An augmented location was removed entirely when it was flagged in more than 25% of the 25 years, corresponding to at least seven annual labels. After this final stage, the augmented sample set contained 332 471 locations, including approximately 25 000 Cropland and 307 000 Non-cropland samples.

Table S1. Principal settings used for augmented-sample construction and annual weak-label screening.

Stage	Component	Parameter or operation	Setting
1	Agreement screening	Reference training samples	FAST-Crop
	Agreement screening	Candidate retention	CatBoost prediction agrees with GLAD-stratified label
	CatBoost	Iterations; depth; learning rate	2000; 8; 0.05
	CatBoost	L2 leaf regularization	6.0
	CatBoost	Loss; evaluation metric	Logloss; F1
	CatBoost	Random seed; thread count	42; 80
	CatBoost	Production task type	CPU
	Spatial filtering	Geographic grid	5°×5°
2	Spectral filtering	Input bands	Blue, Green, Red, NIR, SWIR1, SWIR2
	Spectral filtering	Transformation	$z=\log(1+SR)$
	Spectral filtering	Reference distribution	Six-dimensional Gaussian distribution of transformed FAST-Crop samples
	Spectral filtering	Hard prefilter	$D2 \leq 22.457744$
	Spectral filtering	Sampling method; temperature	Likelihood; 1

Stage	Component	Parameter or operation	Setting
3	Category filtering	Class allocation	Grid-level FAST-Crop Cropland/Non-cropland proportions
	Sample optimization	Candidate batches per iteration	10
	Sample optimization	Batch size	10 000
	Sample optimization	Selection criterion	Highest estimated P_{∞}
	Sample optimization	Target before annual screening	Approximately 340 000
	Sample optimization	Unselected candidates	Returned to candidate pool
	Annual label extension	Temporal assignment	Temporally nearest GLAD Cropland epoch
4	CleanLab	Cross-validation	Five-fold out-of-fold
	CleanLab	Fold construction	StratifiedKFold(n_splits=5, shuffle=False)
	CleanLab	Label-issue detection	Default find_label_issues
	CleanLab	Manually defined threshold	None
	CleanLab	Labels eligible for screening	Augmented point-year labels only
	CleanLab	Flagged point-year	Excluded from corresponding annual classifier
	CleanLab	Location-level removal	Flagged in >25% of years, i.e. at least 7 years

S1.2 Classifier benchmark

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.

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

Classifier	F ₁ 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.

S1.3 Sensitivity of subclass-specific cropland recognition to spatial sample augmentation

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.

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.

S2. Sensitivity of the minimum slope threshold

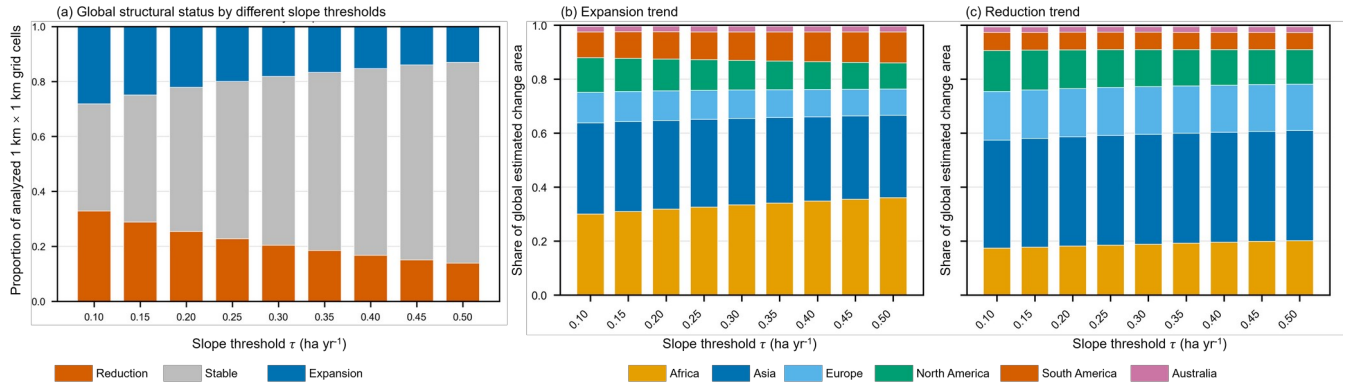


Figure S1. Sensitivity of global structural evolution trend composition and the continental distribution of cropland change to the minimum slope threshold. (a) Proportions of analysed 1 × 1 km grid cells assigned to general expansion ($\beta > \tau$), general reduction ($\beta < -\tau$), and general 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.

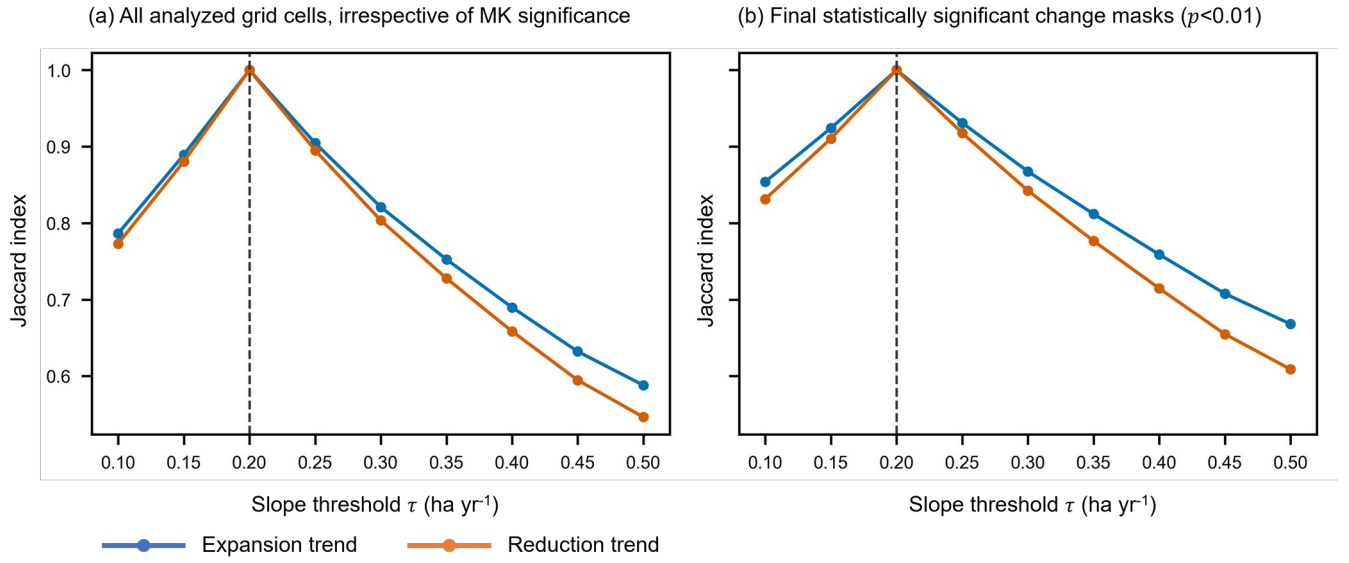


Figure S2. Sensitivity of the spatial expansion and reduction masks to the minimum slope threshold. The conventional Jaccard index, $J = |M_\tau \cap M_{0.2}| / |M_\tau \cup M_{0.2}|$, compares the mask M obtained at each minimum slope threshold τ with the corresponding reference mask at $\tau = 0.2$. 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 $\beta > \tau$, and reduction requires $p < 0.01$ and $\beta < -\tau$. A Jaccard index of 1 indicates identical masks. Panel (b) corresponds to the results of the statistically significant change masks.

S3. Regional evaluation

Table S4. Static thematic accuracy of GACED30 stratified by 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	\	\	\	\	\