remote sensing

MDPI

Article

Quality-Gated Integration of Predicted Fields for Maize–Soybean Mapping with AlphaEarth Foundations

Runze Kang [1,2,3], Ran Bi [2] and Jie Liu [1,3,*]

[1] Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China; 22b903097@stu.hit.edu.cn
[2] China Telecom Artificial Intelligence Technology (Beijing) Co., Ltd., Beijing 100032, China; bir3@chinatelecom.cn
[3] State Key Laboratory of Smart Farm Technologies and Systems, Harbin 150030, China
* Correspondence: jieliu@hit.edu.cn

Highlights

What are the main findings?

- Unsupervised AlphaEarth embeddings achieved high accuracy (macro-F1 0.943) for maize–soybean mapping.
- A quality gate retained a 0.936 macro-F1 while reducing small map patches by 31.0%.

What are the implications of the main findings?

- Global embeddings and predicted field boundaries enable timely crop mapping without local cadastral data.
- Quality-gated parcel refinement effectively balances spatial regularity with the preservation of pixel-level crop information.

Abstract

Timely field-level crop mapping remains difficult where current labels and reliable parcel boundaries are scarce. This study examined whether 2024 AlphaEarth Foundations (AEF) annual embeddings could support maize–soybean mapping through unsupervised clustering followed by label-assisted semantic assignment and whether Fields of the World (FTW) predictions could convert pixel clusters into reliable parcel-level maps. The 64-dimensional AEF features at 10 m resolution were clustered without crop labels in an agricultural region of northern Northeast China. After the cluster solution was fixed, the field observations were used for semantic assignment and spatially blocked assessment. At the 652 route-based reference points collected in this study, direct AEF clustering yielded F1-scores of 0.944 for maize and 0.942 for soybean, with a macro-F1 of 0.943 for binary maize–soybean retrieval. Unconditional parcel assignment increased spatial regularity but reduced agreement with the reference observations. A quality gate based on pixel support, within-field purity, and under-segmentation risk retained a macro-F1 of 0.936 while reducing small patches by 31.0%. Selective use of predicted fields reduced fragmentation while preserving most of the sample-based accuracy of direct AEF clustering, offering a way to control the loss of pixel-level crop information during parcel aggregation.

Keywords: AlphaEarth Foundations; Fields of the World; geospatial foundation model; crop mapping; maize; soybean; unsupervised clustering; field boundary; quality gate

Check for updates

Academic Editors: Abid Ali and Hans-Peter Kaul

Received: 11 September 2026
Revised: 17 October 2026
Accepted: 28 October 2026
Published: 10 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Timely crop-type maps contribute to agricultural statistics, crop-condition monitoring, yield forecasting, rotation analysis, and food-security assessment [1–3]. Their spatial detail reveals field configuration, crop concentration, and local transitions that are absent from aggregate statistics. Seasonal differences in canopy structure, greenness, and moisture make satellite time series well suited to annual crop mapping [4]. In Northeast China, maize and soybean dominate much of the dryland cropping system, and changes in their distribution are relevant to production planning and rotation policy. Existing 10 m products have demonstrated the feasibility of distinguishing these crops at regional scales [5,6]. Annual updates nevertheless remain difficult where current-year labels are sparse and agricultural fields are interspersed with roads, settlements, and shelterbelts.

Most crop maps rely on supervised classifiers trained with observations from the target year [7–9]. Such models require samples that represent variation in crop rotation, weather, management, planting date, and sensor conditions. Dense annual field surveys are costly, and samples collected along a limited route may transfer poorly beyond the surveyed landscape. Evaluation is also affected by spatial autocorrelation: nearby training and validation observations can share environmental and image context, producing optimistic estimates of spatial transfer [10,11]. These constraints motivate methods that learn crop-relevant structure before labels are introduced and reserve spatially separated samples for interpretation and evaluation.

Reusable Earth-observation representations offer an alternative to reconstructing crop features from raw imagery for each region and year. Geospatial foundation models learn features from large multi-sensor archives and provide compact inputs for downstream analysis [12,13]. AlphaEarth Foundations (AEF) combines spatial, temporal, and measurement context in annual 64-dimensional embeddings at a nominal 10 m resolution [14,15]. These embeddings are neither crop labels nor explicit phenological variables; their agricultural value depends on whether crop-related differences remain separable in the feature space. Recent evaluations report promising performance in agricultural tasks while noting limits in spatial transfer, temporal sensitivity, and interpretation [16,17]. Predicted field boundaries provide a different type of information. When polygons correspond to actual management fields with sufficiently consistent crop composition, object-based aggregation can combine pixel evidence, suppress isolated predictions, and produce units suited to agricultural applications [18,19]. Fields of the World (FTW) extends field-boundary predictions to regions where cadastral data are unavailable [20,21]. Automated boundary extraction can nevertheless leave displacement, omission, and over- or under-segmentation errors [22,23]. A displaced boundary or an under-segmented polygon may combine adjacent crops or non-field surfaces. Assigning a single dominant label to such a polygon can overwrite valid minority patterns and propagate local errors across the polygon. Omitted fields reduce coverage when the polygons define the reporting domain, whereas over-segmentation divides a management field into smaller units and can reduce the evidence available for aggregation. AEF describes similarity in land-surface characteristics, whereas FTW proposes the spatial units within which that information might be organized. Their integration therefore requires checking within-field consistency and geometric risk before imposing a uniform label.

Agricultural benchmarking of AEF [16] and supervised tomato mapping with an embedding-based U-Net [17] provide examples of its downstream use. Related approaches include uncertainty-aware crop classification [24] and class-specific conditional random fields for remote-sensing segmentation [25]. The methodological question addressed here is how to combine crop-relevant structure in fixed annual AEF embeddings with predicted field boundaries whose suitability for parcel-wide assignment is uncertain. In a clustering-

based crop workflow, obtaining a feature partition does not by itself determine whether a predicted polygon should only delimit the reporting domain or also replace within-field variation with a single label. We therefore distinguish the thematic information supplied by AEF from the spatial constraints supplied by FTW and examine when parcel-wide assignment is supported by within-polygon evidence.

Using 2024 AEF embeddings, FTW predictions, and maize–soybean observations from northern Northeast China, we examine when predicted fields can regularize a crop map without erasing useful pixel-level variation. We compare direct pixel clustering, geometric masking, majority homogenization, parcel-feature clustering, and quality-gated refinement on a common FTW domain, separating changes in map support from changes in crop labels. The quality gate permits parcel-wide majority assignment only where pixel support, within-field purity, and under-segmentation diagnostics support it; other parcels retain their direct-pixel labels. Crop observations enter after clustering to assign crop names and assess the resulting maps.

2. Materials and Methods

2.1. Study Area

The study covered a rectangular agricultural landscape window centered on Nehe in northwestern Heilongjiang Province, China. Figure 1 shows its location and spatial composition. It extended from 124.3112 to 126.0393°E and from 47.9253 to 48.9556°N, with a projected area of approximately 14,645 km^2. The extent was defined by the overlap between the AlphaEarth Foundations (AEF) export and the Fields of the World (FTW) extraction and does not follow the administrative boundary of Nehe City. All reported map statistics refer to this common extent.

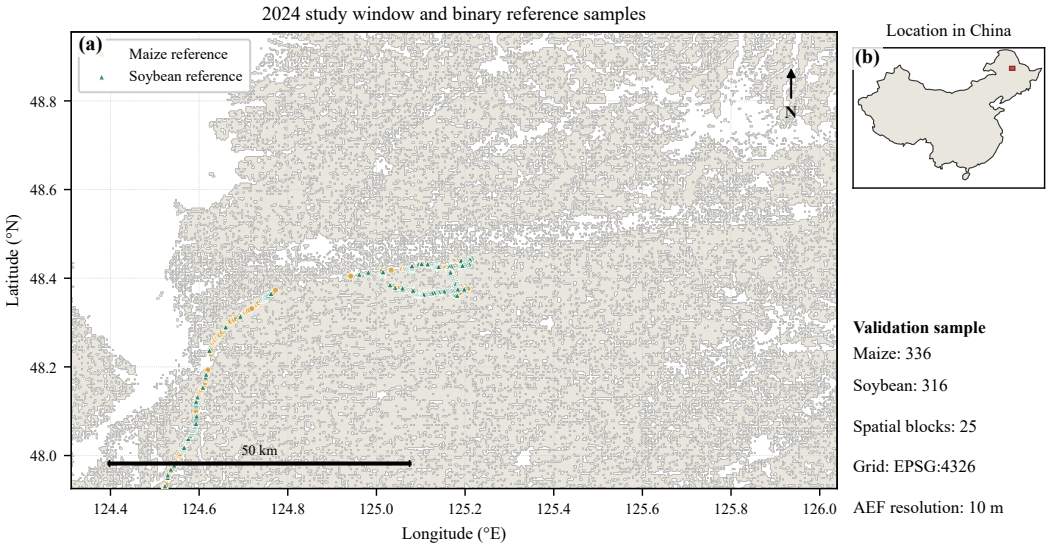

Figure 1. Location and spatial composition of the 2024 analysis window centered on Nehe, Heilongjiang Province. (**a**) The analysis window, showing AlphaEarth Foundations coverage, Fields of the World-predicted field footprints, and the 652 maize–soybean observations in the common assessment set; (**b**) the location of the analysis window within China.

The study window lies in the northern Songnen Plain, an intensively cultivated part of Northeast China. Large mechanized fields are common, and maize and soybean dominate the available 2024 crop observations. Roads, shelterbelts, villages, river corridors, and woodland patches interrupt the cultivated matrix. Adjacent fields often share long boundaries, and predicted segmentation may combine neighboring management units.

This landscape structure is relevant to both pixel-level crop discrimination and parcel-based spatial aggregation.

Rasters and vectors were aligned in WGS 84 coordinates under EPSG:4326, following the source AEF grid and FTW polygons. Metric geometry was calculated in WGS 84/UTM zone 51N under EPSG:32651. The study-area map shows AEF coverage, FTW footprints, and the maize and soybean observations retained for assessment. Because the window is not an administrative unit and the observations were not selected using probability sampling, the analysis does not provide administrative crop-area estimates.

2.2. Datasets

2.2.1. AlphaEarth Satellite Embeddings

Remote-sensing features were obtained from the 2024 Google Satellite Embedding V1 Annual product, catalog identifier GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL. The product provides 64-dimensional annual embeddings at a nominal spatial resolution of 10 m [14,15]. Unlike reflectance observations from individual acquisition dates, AEF represents the spatial and temporal context learned from multi-source Earth observations over the year. This compact annual representation is well suited to landscapes in which crop differences emerge through seasonal development as well as surface condition, although its individual dimensions do not correspond directly to named biophysical variables.

The tiles covering the study window were assembled on the native WGS 84 grid, which was retained as the common spatial frame for all subsequent analyses. Only pixels with complete values across the 64 embedding dimensions were used. The released embeddings were analyzed without task-specific fine-tuning, and no reflectance bands, vegetation indices, weather variables, crop calendars, or external phenological metrics were added. The resulting feature set therefore isolated the information contained in the annual AEF representation.

2.2.2. Fields of the World Field Boundaries

Field geometry was obtained from the 2024 global Fields of the World predictions for the Heilongjiang partition, distributed through Source Cooperative [20,21]. FTW provides a spatially consistent set of field-like polygons for regions where cadastral or locally surveyed boundaries are unavailable. The boundaries were inferred from Sentinel-2 imagery and polygonized from the provider's field-probability output at the nominal 0.5 threshold. As an automatically predicted product, FTW represents probable field footprints rather than legal parcels or verified single-crop management units; boundary displacement, omission, and over- or under-segmentation may remain.

Polygons intersecting the study window were extracted and aligned with the AEF grid. After basic geometry checks, they served two distinct purposes: delineating the mapped field domain and defining spatial units for aggregation of AEF features. Their suitability for geometric masking was considered separately from their suitability for parcel-wide label assignment. FTW was therefore not treated as an independent cropland label or as field-boundary ground truth; attribute refinement depended instead on pixel support and within-polygon consistency described in Sections 2.3 and 2.5.3.

2.2.3. Crop Reference Samples

Crop reference information came from geolocated field surveys conducted in 2024. Each record identifies the crop observed at a surveyed location. The observations are concentrated along field-survey routes and were not selected through probability sampling; they provide local reference information for cluster interpretation and point-based validation, but they do not support direct estimation of regional crop proportions.

Only records temporally matched to the 2024 AEF and FTW products were considered. Survey locations were harmonized to WGS 84 after coordinate plausibility and spatial-consistency checks, and crop names were standardized to a common vocabulary. Records with missing or ambiguous crop identity and locations without valid spatial data were excluded. The analysis retained 652 maize–soybean observations, including 336 maize and 316 soybean records, that had comparable predictions across the mapping strategies.

The route-based distribution produced local concentrations of observations. Spatial blocks were defined by a regular angular grid in WGS 84, with a primary cell width of $s = 0.05^\circ$ in both longitude and latitude. The grid origin was $(\lambda_0, \phi_0) = (124.521388^\circ\text{E}, 47.925967^\circ\text{N})$. Block indices were $b_x = \lfloor(\lambda - \lambda_0)/s\rfloor$ and $b_y = \lfloor(\phi - \phi_0)/s\rfloor$. The 652 reference observations occupied 25 cells, with 1–83 observations per occupied cell. At the median reference latitude of 48.3733°N, a cell spans approximately 3.70 km east–west and 5.56 km north–south; these are angular grid cells rather than equal-distance squares or administrative units.

The occupied blocks were assigned to five folds using deterministic GroupKFold partitioning, which balances observation counts subject to keeping each block intact. The primary held-out folds contained 129–133 observations. This grouping limits within-block sharing between training and test sets [10], but no spatial buffer was imposed between blocks in different folds. The partition was applied after the unsupervised models and their settings had been fixed and was used only for cluster interpretation and out-of-fold assessment.

Workflow overview. Figure 2 summarizes the primary workflow using schematic inputs, parcel-label patterns, and spatial folds. The sequence separates label-free feature construction and spatial refinement from subsequent label-assisted crop naming and assessment; the four comparison strategies are shown alongside the quality-gated path.

2.3. Data Preprocessing and Field Characterization

Empty FTW geometries were removed, invalid polygons were repaired, and multipart features were decomposed before raster overlay. Positive-area overlap was distinguished from boundary contact so that neighboring fields sharing an edge were represented as adjacent units rather than overlapping areas. No independent cropland mask was applied, and the processed polygons retained their interpretation as predicted field footprints rather than independent evidence of cropland presence.

For each polygon p , area A_p and perimeter P_p were calculated in EPSG:32651. Compactness was defined as

$$C_p = \frac{4\pi A_p}{P_p^2}, \tag{1}$$

where values approach one for compact shapes and decline for elongated or irregular polygons. Exact-touch adjacency pairs and connected components were derived from the vector topology. Positive-area overlap, adjacency degree, and component size were recorded separately because contiguous fields may legitimately share a boundary even when no two polygon interiors overlap.

Converting raster predictions into vector units is a nontrivial operation that has been addressed in other domains through adaptive grid projection [26]. In the present study, the alignment of AEF pixels with FTW polygons relied on fractional subpixel coverage rather than adaptive projection, but the general problem of how to attribute grid-based information to irregular vector geometries is shared.

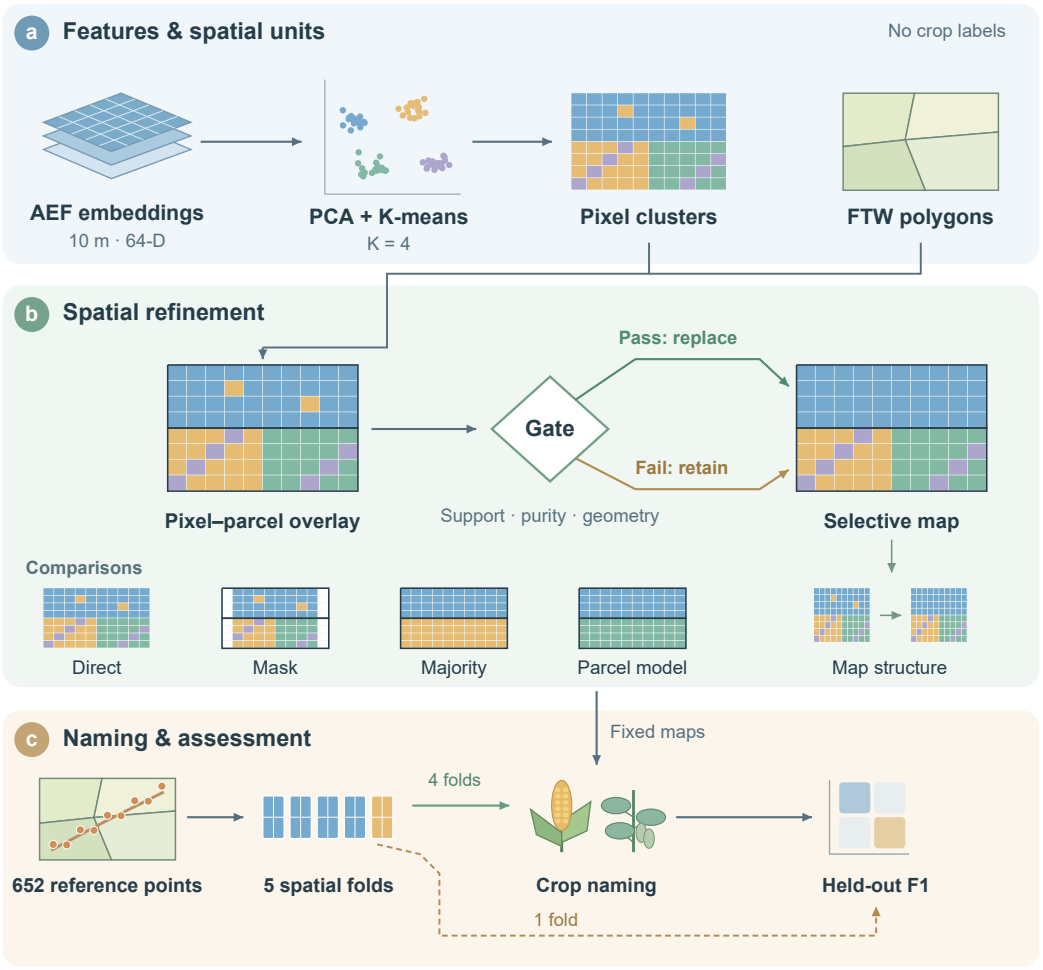

Figure 2. Pictorial overview of the primary $K = 4$ workflow. (**a**) AEF embeddings supply pixel clusters, and FTW supplies predicted spatial units. (**b**) The gate requires $n_p \geq 10$, $\rho_p \geq 0.90$, and no under-segmentation flag. Accepted polygons receive the dominant pixel label; other polygons retain their pixel labels. Thumbnails summarize the direct-pixel, mask, majority, and parcel-feature comparisons; parcel-feature clustering is independent of the gate. (**c**) Four training folds name fixed candidate clusters 0 and 1 as maize or soybean; one held-out fold supplies labels for assessment. Clusters 2 and 3 remain rejected. Map structure and crop accuracy are assessed separately, with strategy comparisons on the common FTW domain. Arrows indicate the direction of data processing and assessment; the pass and fail branches indicate parcel-wide replacement and retention of direct-pixel labels, respectively. All icons, point clouds, and raster patterns are schematic; colors before crop naming denote clusters rather than validated crop classes.

Each FTW polygon was overlaid with the AEF grid. Fractional pixel coverage was approximated using a regular 4×4 grid of subpixel centers, giving 16 inclusion tests per native AEF pixel. If $w_{pi} \in [0, 1]$ denotes the resulting fractional support of pixel i in parcel p and x_{id} is embedding dimension d, the parcel-weighted mean was

$$\bar{x}_{pd} = \frac{\sum_{i \in p} w_{pi} x_{id}}{\sum_{i \in p} w_{pi}}, \qquad d = 1, \ldots, 64. \tag{2}$$

The within-parcel weighted standard deviation was

$$s_{pd} = \left[\frac{\sum_{i \in p} w_{pi} \left(x_{id} - \bar{x}_{pd} \right)^2}{\sum_{i \in p} w_{pi}} \right]^{1/2}. \tag{3}$$

A native pixel contributed to the support count n_p only when its center was inside the polygon and all 64 bands were valid. Parcel-feature modeling required at least three valid pixels, while polygons below this threshold remained available for geometric masking.

The 64 weighted means constituted the parcel clustering feature vector. The 64 weighted standard deviations were retained as heterogeneity diagnostics rather than appended to the primary clustering input. For a given direct-pixel cluster map, parcel purity was defined as

$$\rho_p = \max_k \frac{n_{pk}}{\sum_j n_{pj}}, \tag{4}$$

where n_{pk} is the number of valid pixel centers assigned to cluster k within parcel p. Parcel entropy was calculated as $H_p = -\sum_k q_{pk} \log(q_{pk})$, where $q_{pk} = n_{pk} / \sum_j n_{pj}$ and zero terms were omitted. Purity and entropy describe label consistency, whereas the mean of s_{pd} over 64 dimensions describes feature-space heterogeneity.

Potential under-segmentation was screened without crop labels. A polygon was flagged if its area was at or above the 99th percentile and its mean AEF standard deviation was at or above the 90th percentile; if its area was at or above the 95th percentile and its compactness was at or below the 5th percentile; or if its area exceeded 500 ha. These criteria identified unusually large, heterogeneous, or elongated polygons for which parcel-wide assignment warranted caution. A flag indicated potential under-segmentation rather than a confirmed boundary error.

2.4. Pixel-Level Clustering of AlphaEarth Embeddings

Pixel-level clustering used a spatially distributed sample of 250,000 valid AEF pixels. Systematic sampling across the study window preceded random subsampling so that densely represented local areas did not dominate model fitting.

Each embedding dimension was standardized using the sample mean and standard deviation. Principal component analysis (PCA) retained the minimum number of components needed to explain at least 95% of cumulative variance [27]; 21 components explained 95.30%. Two-dimensional PCA projections were used for visualization only and did not inform crop labels.

K-means models were evaluated for $K = 2, \ldots, 12$, with 50 initializations for each K [28]. Candidate solutions were compared using the silhouette coefficient, Calinski–Harabasz index, and Davies–Bouldin index [29–31]. Higher silhouette and Calinski–Harabasz values and lower Davies–Bouldin values indicate more compact, separated partitions. The three indices did not identify a uniformly dominant pixel solution: $K = 4$ and $K = 5$ formed the leading candidate range, with the former favored by the Davies–Bouldin index and the latter by the silhouette and Calinski–Harabasz indices. We retained the more parsimonious $K = 4$ solution as the primary pixel model. This choice also provided the same partition cardinality as the independently selected $K = 4$ parcel-feature model in Section 2.5.2, allowing the pixel–parcel experiment to isolate changes in spatial support and feature aggregation. The $K = 5$ solution was evaluated separately as a cluster-count sensitivity analysis.

The fixed $K = 4$ model was applied to all valid AEF pixels using the scaling and PCA parameters estimated from the 250,000-pixel sample. This raster served as the baseline for parcel-integration comparisons. Crop observations were excluded from sampling, standardization, PCA, candidate-K evaluation, and full-scene prediction.

2.5. Pixel–Parcel Integration Strategies

Five strategies were evaluated on the same AEF grid to separate three roles of predicted boundaries: restricting map support, modifying existing pixel labels, and modeling aggregated parcel features. Table 1 defines their inputs and outputs. Direct pixel clustering

provided the common baseline. All outputs retained the AEF projection, resolution, alignment, and NoData convention. Cluster identifiers had no crop meaning before the semantic assignment described in Section 2.6.

Table 1. Definition of the five mapping strategies. Reclustering denotes an additional clustering model fitted after the direct-pixel baseline; FTW-derived attributes are parcel statistics calculated from FTW geometry.

Strategy	Input	Spatial Support	Recluster	Changes Pixel Labels	Uses FTW-Derived Attributes	Output
Direct pixel clustering	Standardized, PCA-transformed AEF pixels	All valid AEF pixels	No; baseline fit	No; creates baseline	No	Full-extent map; common-domain view for parcel comparisons
Geometric field masking	Direct labels and FTW geometry	Valid pixels inside FTW footprints	No	No	No; geometry only	Clipped direct-pixel map
Hard parcel homogenization	Direct labels grouped by eligible FTW polygons	Valid pixels inside eligible FTW footprints	No	Yes; all eligible fields	Yes; dominant internal cluster	Uniform majority label per field
Parcel-level feature clustering	Area-weighted 64-D AEF field means and FTW polygons	Valid pixels in fields with sufficient AEF support	Yes	Yes; all eligible fields	Yes; aggregated field features	Uniform independently clustered label per field
Quality-gated refinement	Direct labels, support, purity, and geometry-risk metrics	Geometry-checked FTW footprints	No	Selectively, where $G_p = 1$	Yes; gate criteria	Direct-majority label in accepted fields; direct labels elsewhere

2.5.1. Pixel Baseline and Geometric Masking

Direct pixel clustering. The fixed $K = 4$ pixel model was applied to every valid AEF pixel. FTW polygons were not used for feature extraction, model fitting, cluster assignment, or spatial restriction, producing the full-domain Direct AEF map. For comparisons with parcel-based strategies, the same labels were also represented on the common FTW domain; this derived view is referred to as the common-domain Direct AEF baseline and does not involve any label reassignment.

Geometric field masking. All FTW polygons that passed geometry checks were rasterized on the AEF grid to form a binary mask. Direct-pixel labels within the mask were retained, and pixels outside the predicted footprints were set to NoData. Geometric masking therefore changed map support without changing retained labels. The full-domain Direct AEF map was used to quantify coverage loss, whereas the common-domain Direct AEF view was used as the reference for within-footprint attribute comparisons.

2.5.2. Unconditional Parcel Integration

Majority-vote homogenization. Each eligible parcel received its most frequent direct-pixel cluster, which was then assigned to all valid pixels within the polygon without a purity or heterogeneity criterion. This operation modified the direct result without fitting a parcel-level model.

Parcel-feature clustering. For parcel-level modeling, the 64 area-weighted AEF means from 125,737 eligible polygons were standardized. PCA retained 19 components, which

explained 95.03% of cumulative variance. K-means and Gaussian mixture models (GMMs) were compared for $K = 2, \ldots, 12$ [28,32]. K-means used 50 initializations. For each GMM cluster count, spherical, diagonal, tied, and full covariance structures were fitted, and the solution with the lowest Bayesian information criterion was retained.

Silhouette values were calculated from 5000 parcels, while the Calinski–Harabasz and Davies–Bouldin indices used all eligible parcels. The mean rank of the three indices selected K-means with $K = 4$. Stability was assessed over 100 independent K-means fits, with pairwise adjusted Rand index (ARI) calculated from a common sample of 10,000 parcels [33]. The retained solution combined a low within-cluster objective with high agreement across fits.

Parcel-cluster identifiers were aligned with the four direct-pixel identifiers using a 4×4 spatial-overlap table and Hungarian assignment. The aligned parcel label was then assigned to every valid pixel in the polygon. This strategy fitted a new partition to aggregated features, whereas majority-vote homogenization retained the direct-pixel partition and changed only its spatial assignment.

2.5.3. Quality-Gated Parcel Refinement

Quality-gated refinement used all geometry-checked polygons for masking but restricted parcel-wide label replacement to polygons that met predefined quality criteria.

The primary gate defines an empirical operating point between parcel-level regularization and agreement with the reference observations. Alternative thresholds were examined in the sensitivity analysis, and paired block bootstrap quantified the accuracy difference from direct AEF.

A parcel passed the gate when it had sufficient pixel support, high direct-label purity, and no under-segmentation flag:

$$G_p = \mathbb{I}\left(n_p \geq 10,\ \rho_p \geq \tau,\ p \notin \mathcal{U}\right), \tag{5}$$

where $\tau = 0.90$ was the primary purity threshold and $\mathcal{U}$ is the under-segmentation suspect set. The purity criterion requires at least 90% of valid pixels in an accepted parcel to share the dominant direct-cluster label, so parcel-wide replacement can change at most 10% of its valid pixel labels. The ten-pixel minimum avoids assigning a parcel-wide label from very small pixel counts, while the under-segmentation screen flags potentially unreliable spatial units. These cutoffs are empirical settings, not calibrated guarantees of crop accuracy or boundary correctness. Alternative settings were evaluated through the sensitivity analysis in Section 2.8.

For $G_p = 1$, the dominant direct-pixel label replaced the pixel labels within parcel p; otherwise, the original direct-pixel labels were retained. A parcel could therefore contribute to geometric masking without imposing a uniform attribute. Parcel-feature clustering remained an independent comparison strategy and did not enter the gate.

2.6. Cluster Interpretation and Crop Label Assignment

Dimensionality reduction and clustering were unsupervised; crop naming and accuracy assessment subsequently used reference labels.

The AEF and parcel models, preprocessing transformations, cluster count, and gate definition were fixed before crop naming. For the primary $K = 4$ analysis, clusters 0 and 1 were treated as the fixed candidate target-crop clusters, whereas clusters 2 and 3 remained semantically unassigned. The same five spatial folds were used for each mapping strategy, and each spatial block belonged entirely to one fold.

For each held-out fold f, the reference labels in the other four folds were used to compare the two possible one-to-one assignments of the candidate clusters to maize and

soybean. Let $n_{kc}^{(-f)}$ denote the number of training observations of crop c in cluster k. The two agreement scores were

$$S_{\text{id}}^{(-f)} = n_{0,\text{maize}}^{(-f)} + n_{1,\text{soybean}}^{(-f)}, \qquad S_{\text{swap}}^{(-f)} = n_{0,\text{soybean}}^{(-f)} + n_{1,\text{maize}}^{(-f)}.$$

When $S_{\text{id}}^{(-f)} \geq S_{\text{swap}}^{(-f)}$, cluster 0 was assigned to maize and cluster 1 to soybean; otherwise, the assignments were reversed. The identity assignment was the deterministic tie-breaking rule. The resulting mapping was applied to the held-out fold without using its crop labels, yielding one out-of-fold prediction per observation. Assignment was performed separately for each mapping strategy using its own cluster predictions.

The candidate set was held fixed across folds rather than reselected from each training subset. Clusters 2 and 3 were rejected by this fixed rule, not by a fold-specific support or dominance threshold. A rejected prediction at a maize or soybean observation was counted as a failure to retrieve that crop. Rejected clusters were not assigned other land-cover names, and visual inspection, vegetation indices, and crop-calendar information were not used to revise the crop assignments.

The supplementary $K = 5$ analysis used a different, many-to-one interpretation rule: within each training fold, a cluster was assigned to its dominant crop only if it contained at least 10 target-crop observations and the dominant-crop proportion was at least 0.70; otherwise, it remained unassigned. Several clusters could therefore receive the same crop name. These thresholds were specific to the supplementary interpretation and were not applied in the primary $K = 4$ analysis. The $K = 4$–$K = 5$ comparison therefore reflects both the cluster partition and the interpretation rule.

2.7. Accuracy Assessment and Statistical Analysis

All strategies were evaluated with the same reference observations and spatial folds. Accuracy and paired comparisons used the common FTW domain; at these locations, the common-domain Direct AEF labels are identical to the full-domain Direct AEF labels. A masked-out observation was treated as a rejected prediction. Majority-vote homogenization and parcel-feature clustering were assessed separately because the former reassigns direct-pixel labels and the latter fits a new model to aggregated parcel features.

For each crop c, precision, recall, F1-score, and intersection over union (IoU) were calculated as

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c}, \qquad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c}, \tag{6}$$

$$\text{F1}_c = \frac{2\,\text{Precision}_c\text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}, \qquad \text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}. \tag{7}$$

Macro-F1 and macro-IoU were the unweighted means of the two crop-specific values, and balanced accuracy was the mean crop recall. Target coverage was the proportion of observations predicted as maize or soybean. Rejected predictions were counted as false negatives for the reference crop but not as false positives for the other target crop. Overall accuracy for an all-crop population and area-adjusted accuracy were not estimated because the reference classes and sampling design do not support those quantities [34,35].

Uncertainty was estimated from 2000 block-bootstrap replicates. The 25 spatial blocks were sampled with replacement, and all observations in each selected block were retained. The 2.5th and 97.5th percentiles of the bootstrap distribution defined the 95% confidence interval.

Paired differences were calculated on the common sample set by drawing the same blocks for both products in each bootstrap replicate and subtracting the direct-pixel metric from the comparison metric. Correct and incorrect disagreements were also summarized

with the exact two-sided McNemar test [36]. These comparisons describe differences at the observed target-crop locations and do not provide probability-sample inference for the mapped region.

Map structure was assessed independently of crop accuracy using changed-pixel counts, valid map support, parcel purity, parcel entropy, ARI, and normalized mutual information (NMI). Purity and entropy were summarized as both unweighted and valid-pixel-weighted means. Small patches were defined as eight-neighbor connected components with fewer than 10 pixels, approximately 0.1 ha at the nominal resolution. ARI and NMI were calculated over the common valid-pixel domain. The cartographic boundary-cut version was excluded from these comparisons.

2.8. Sensitivity Analysis

Gate sensitivity was evaluated for purity thresholds $\tau \in \{0.80, 0.85, 0.90, 0.95\}$ and minimum valid-pixel counts $n_{\min} \in \{3, 5, 10, 20\}$. Feature heterogeneity was defined as $h_p = 64^{-1} \sum_{d=1}^{64} s_{pd}$. Additional variants retained parcels below the 75th, 85th, 90th, or 95th percentile of h_p, together with a variant without an explicit heterogeneity threshold. The threshold sweep recorded accepted parcels, accepted-pixel support, label changes, target coverage, crop-specific and macro accuracy, and parcel purity and entropy. Small-patch abundance was evaluated for the primary gate and the comparison strategies, not for each threshold combination.

Sensitivity of the unsupervised structure was examined across $K = 2$–12, K-means and GMM parcel models, repeated initializations, and spatial subsamples. Feature variants included standardized 64-dimensional embeddings without PCA and PCA representations retaining 90%, 95%, or 99% of cumulative variance. The primary $K = 4$ pixel solution was compared with the alternative $K = 5$ partition identified by the internal diagnostics. ARI and NMI summarized agreement among partitions, and crop-level differences used the same spatial folds and paired block bootstrap as the primary analysis.

The primary pixel solution used $K = 4$, and the parcel-feature comparison retained 95% of PCA variance. The gate used $n_p \geq 10$, $\rho_p \geq 0.90$, and exclusion of polygons flagged for under-segmentation. These settings were fixed before reference labels were used for assessment; other combinations were treated as sensitivity analyses rather than models selected from validation accuracy.

Spatial-block size was additionally examined at $s = 0.025^\circ$, 0.05°, and 0.10°, keeping the original grid origin and all 652 observations fixed. At each scale, five group folds were reconstructed and crop naming was relearned within each training subset from the stored cluster predictions. Clustering, gate parameters, and the candidate cluster set were unchanged. Accuracy intervals used 2000 bootstrap replicates of the occupied blocks at that scale; the quality-gated-minus-direct comparison resampled the same blocks for both strategies. The analysis evaluates block-size sensitivity conditional on the same observations and candidate clusters.

3. Results

3.1. Data Coverage and Field Characteristics

The 2024 AEF mosaic contained 220,629,153 valid pixels within the analysis window. The FTW layer contained 140,971 predicted field footprints, of which 125,737 had at least three valid AEF pixels and were included in parcel-feature modeling. FTW masking retained 133,907,582 pixels, or 60.69% of the valid AEF domain. The common reference set comprised 336 maize and 316 soybean observations distributed among 25 spatial blocks. Figure 3 and Table 2 summarize data coverage and field characteristics.

FTW field area and AEF pixel support were strongly right-skewed. Median field area was 0.312 ha, with an interquartile range of 0.060–1.779 ha, and the 95th percentile was 33.41 ha. The median valid-pixel count was 47, with an interquartile range of 9–266. Many small fields had uniform direct-pixel labels. Across all fields, however, pixel-weighted mean purity was 0.807 and pixel-weighted entropy was 0.387.

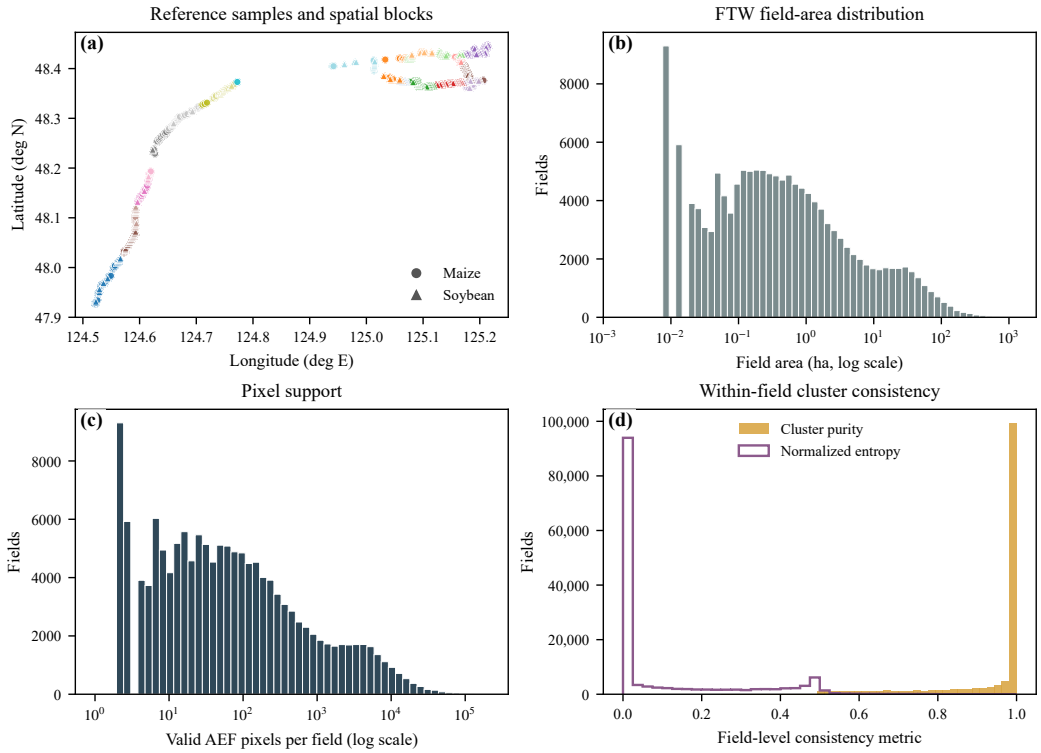

Figure 3. Reference-sample coverage and FTW field characteristics. (**a**) Maize and soybean observations in 25 spatial blocks assigned to five cross-validation folds; distributions of (**b**) predicted field area, (**c**) valid AEF pixel support, and (**d**) within-field direct-cluster consistency.

Table 2. Data coverage and field characteristics.

Quantity	Value	Scope
Valid AEF pixels	220,629,153	Full analysis window
FTW-contained AEF pixels	133,907,582	60.69% of valid AEF
Geometry-checked FTW fields	140,971	All available for masking
Fields eligible for AEF modeling	125,737	At least three valid pixels
Field area (ha)	0.312	IQR 0.060–1.779
Valid pixels per field	47	IQR 9–266
Reference observations	652	336 maize; 316 soybean

3.2. AlphaEarth Cluster Structure and Stability

PCA retained 21 components for direct pixels and 19 for parcel means, explaining 95.30% and 95.03% of cumulative variance, respectively. At pixel scale, the silhouette values were 0.292 for $K = 5$ and 0.273 for $K = 4$. The corresponding Calinski–Harabasz values were 69,368 and 67,392. By contrast, $K = 4$ produced the lower Davies–Bouldin value, at 1.182 compared with 1.342 for $K = 5$. The diagnostics therefore supported a candidate range rather than a single uniformly superior cluster count. The primary analysis retained the more parsimonious $K = 4$ partition, while $K = 5$ was carried forward as a sensitivity case. The parcel analysis independently selected K-means with $K = 4$. Among 100 parcel-model fits, 98 had a mean pairwise ARI above 0.98, and the retained solution had a mean pairwise

ARI of 0.9884. Across 20 spatial subsamples, ARI relative to the primary partition averaged 0.9915 and ranged from 0.9795 to 0.9964. Figure 4 and Table 3 present the cluster diagnostics and stability results.

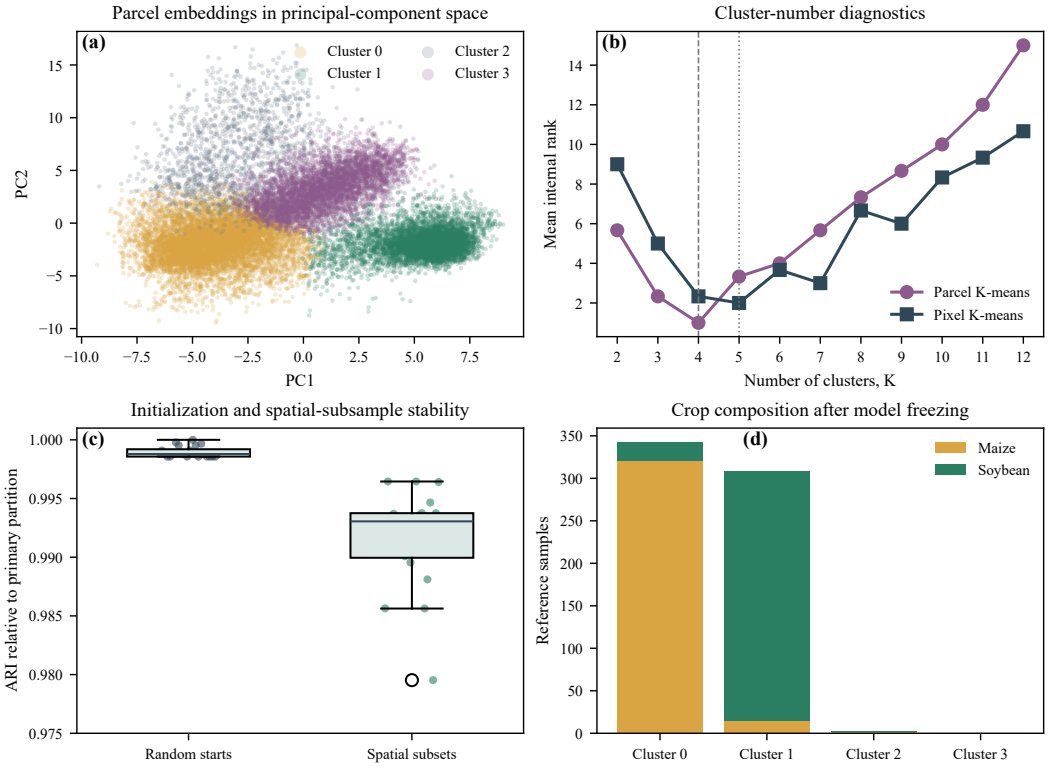

Figure 4. AEF cluster structure and stability across analysis scales. (**a**) Parcel-mean embeddings in the first two principal components; (**b**) mean internal-index ranks for parcel and pixel K-means across $K = 2, \ldots, 12$ (lower is better); the dashed and dotted vertical lines mark $K = 4$ and $K = 5$, respectively; (**c**) parcel-cluster agreement under random initialization and spatial subsampling, where colored points show individual runs or subsamples, boxes span the interquartile range, center lines show medians, whiskers extend to the most extreme values within 1.5 interquartile ranges, and open circles mark values beyond the whiskers; and (**d**) maize–soybean reference composition of the fixed direct-pixel clusters. Panels (**a**,**c**) present parcel-scale results, and panels (**b**,**d**) present pixel-scale results.

Table 3. Cluster selection and stability summary.

Component	Result	Criterion
Pixel PCA	21 components	95.30% variance
Parcel PCA	19 components	95.03% variance
Pixel candidate range	K-means, $K = 4$–5	Internal indices disagreed
Primary pixel model	K-means, $K = 4$	Parsimony and controlled scale comparison
Alternative pixel model	K-means, $K = 5$	Cluster-count sensitivity
Parcel model	K-means, $K = 4$	Best parcel internal rank
Repeated initialization	98/100 runs	Mean pairwise ARI > 0.98
Spatial subsampling	0.9915	Mean ARI relative to primary

Across the full reference set, direct cluster 0 contained 320 maize and 22 soybean observations, while cluster 1 contained 14 maize and 294 soybean observations. These totals describe the fixed partition; foldwise assignments were determined from the training subsets only. Table A1 reports the training support, crop composition, and assigned meaning of all four clusters in each fold. Cluster 0 was mapped to maize and cluster 1 to soybean in all five folds, with no tied assignment scores. Training-fold maize proportions in cluster 0 ranged

from 92.54% to 94.49%, and soybean proportions in cluster 1 ranged from 94.80% to 96.39%. The selected assignment matched 485–497 training observations per fold, compared with 24–33 for the reversed assignment.

The training subsets contained only one or two target-crop observations in cluster 2 and none in cluster 3; both clusters remained rejected under the fixed candidate-set rule. The other four mapping strategies also retained the cluster 0-to-maize and cluster 1-to-soybean assignment in all five folds. Crop naming was thus stable across the tested folds for each fixed partition.

3.3. Spatial Patterns of Direct Pixel Clustering

Figure 5 shows that direct AEF clustering produced extensive, field-oriented regions of clusters 0 and 1 without using FTW attributes. The full-domain map and its common-domain representation have identical labels wherever FTW footprints are retained; only the latter has pixels outside the predicted footprints set to NoData. Mean field purity was 0.938 when fields were weighted equally and 0.807 when weighted by valid pixel count. The common-domain direct map contained 241,029 connected components smaller than 10 pixels, representing 0.563% of field-contained pixels. Mixed labels occurred most often near field edges and within low-purity footprints.

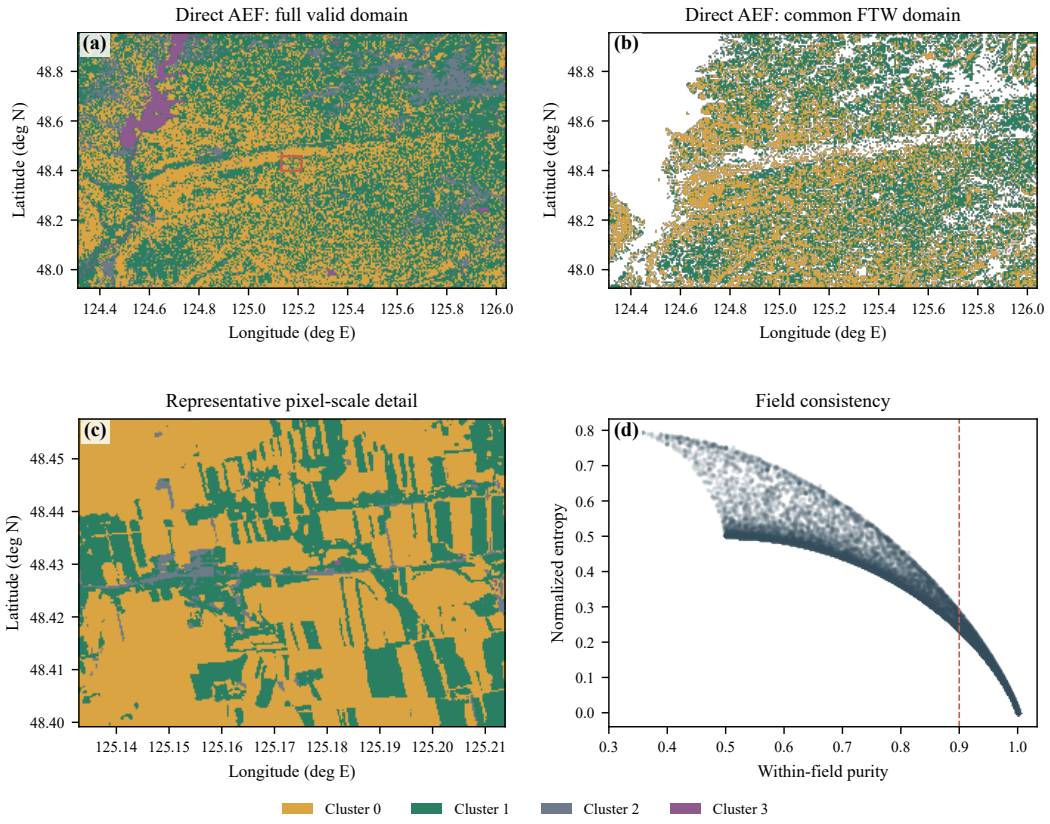

Figure 5. Spatial structure of direct AEF clustering. (**a**) Full-domain Direct AEF map; (**b**) the same labels restricted to the common FTW domain; (**c**) representative pixel-scale detail; and (**d**) relationship between within-field cluster purity and normalized entropy. The red rectangle in panel (**a**) marks the detail window shown in panel (**c**); the red dashed line in panel (**d**) marks the primary purity threshold of 0.90. The two maps differ only in spatial support, not in retained pixel labels.

3.4. Effects of Parcel Integration

Geometric masking removed 86,721,571 valid AEF pixels, equivalent to 39.31% of the complete AEF domain. This difference is visible when the full-domain and common-

domain Direct AEF maps are compared in Figure 5. Masking did not change any retained cluster label, so ARI and NMI relative to direct clustering were both 1.0 on the common field domain. All 652 reference observations remained inside the analysis mask, and their predicted target classes were unchanged.

3.4.1. Unconditional Parcel Integration

Hard parcel homogenization. Majority-vote homogenization changed 25,812,555 pixels, equivalent to 19.28% of field-contained pixels. ARI and NMI relative to direct clustering were 0.387 and 0.309, respectively. Field entropy was zero after homogenization, and the number of small patches decreased from 241,029 to 29,194. These spatial changes are shown in Figure 6.

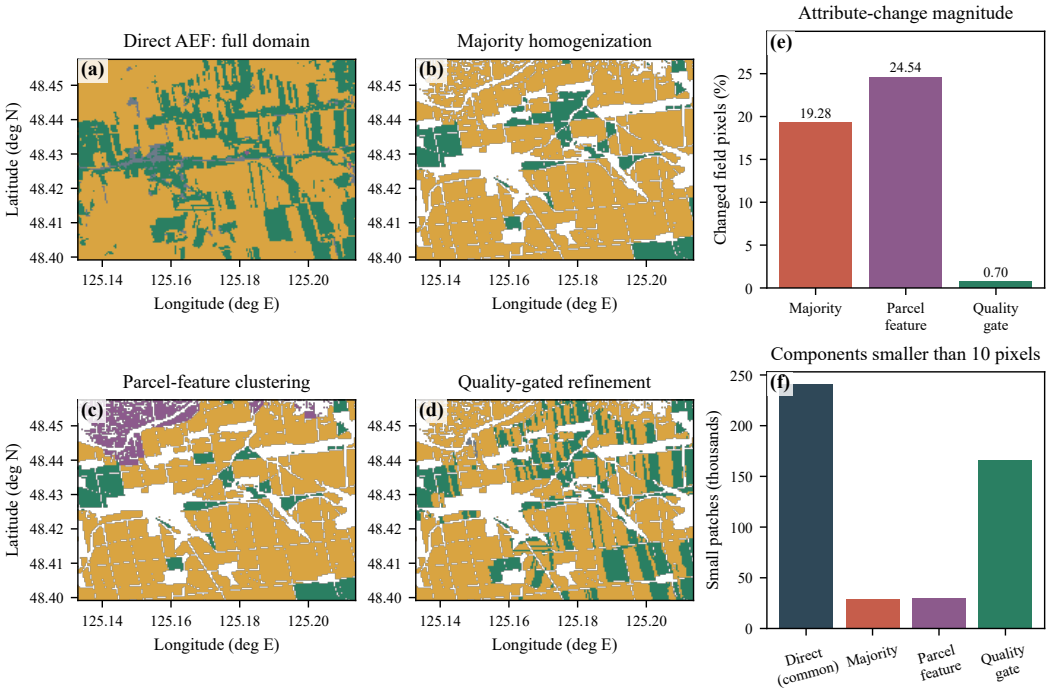

Figure 6. Effects of parcel integration in a common field-scale window: (**a**) full-domain Direct AEF, (**b**) majority homogenization, (**c**) parcel-feature clustering, and (**d**) quality-gated refinement. Panels (**e**,**f**) report label changes and small-patch abundance on the common FTW domain; the direct bar in panel (**f**) therefore refers to common-domain Direct AEF labels. Map colors indicate cluster identifiers: ochre, cluster 0; green, cluster 1; gray-blue, cluster 2; and purple, cluster 3. White areas denote pixels outside the displayed FTW support.

Parcel-level feature clustering. Clustering the parcel-mean embeddings changed 24.54% of field-contained pixels. Agreement with the direct partition was lower than for majority homogenization, with an ARI of 0.316 and NMI of 0.286. The small-patch count fell to 30,246. Both unconditional strategies reduced within-parcel fragmentation, while parcel-feature clustering differed more from the direct AEF partition than majority homogenization.

3.4.2. Quality-Gated Parcel Refinement

The primary gate accepted 78,891 fields, representing 62.74% of FTW fields with sufficient AEF support and covering 50,284,404 pixels. Labels changed for 942,040 pixels, or 0.704% of the FTW domain.

Quality-gated refinement increased pixel-weighted field purity from 0.807 to 0.814, reduced entropy from 0.387 to 0.357, and reduced the number of small patches by 31.0%. Its ARI and NMI relative to direct AEF were 0.974 and 0.942. The changed-pixel fraction

was lower, and agreement with the direct partition was higher, than for either unconditional parcel strategy. The 31.0% reduction was measured at the primary gate settings; its sensitivity to alternative thresholds was not evaluated.

Figure 6 uses the full-domain Direct AEF map in panel a so that the spatial restriction imposed by FTW remains visible. The numerical comparisons in panels e and f are calculated only on the common FTW domain and therefore isolate attribute changes from differences in map support. Table 4 summarizes these spatial effects for all four attribute strategies.

Table 4. Spatial effects of parcel integration. Change and agreement are relative to direct AEF clustering on the common FTW domain.

Method	Changed (%)	ARI	NMI	Entropy	Small Patches
Direct AEF	0.000	1.000	1.000	0.387	241,029
Majority homogenization	19.276	0.387	0.309	0.000	29,194
Parcel-feature clustering	24.543	0.316	0.286	0.000	30,246
Quality-gated refinement	0.704	0.974	0.942	0.357	166,210

3.5. Classification Accuracy and Paired Comparisons

At the common reference locations, the full-domain and common-domain Direct AEF predictions were identical. Direct AEF clustering yielded F1-scores of 0.944 for maize and 0.942 for soybean, with a macro-F1 of 0.943. Under majority homogenization, maize and soybean F1 declined to 0.815 and 0.768. Parcel-feature clustering yielded F1-scores of 0.799 for maize and 0.737 for soybean, with a macro-F1 of 0.768. Soybean recall declined from 0.930 for direct AEF to 0.703 under majority homogenization and 0.661 under parcel-feature clustering.

Quality-gated refinement yielded F1-scores of 0.939 for maize and 0.934 for soybean and a macro-F1 of 0.936. One target observation was rejected. Cross-crop errors were fewer than under either unconditional parcel treatment. Figure 7 and Table 5 provide the crop-specific results.

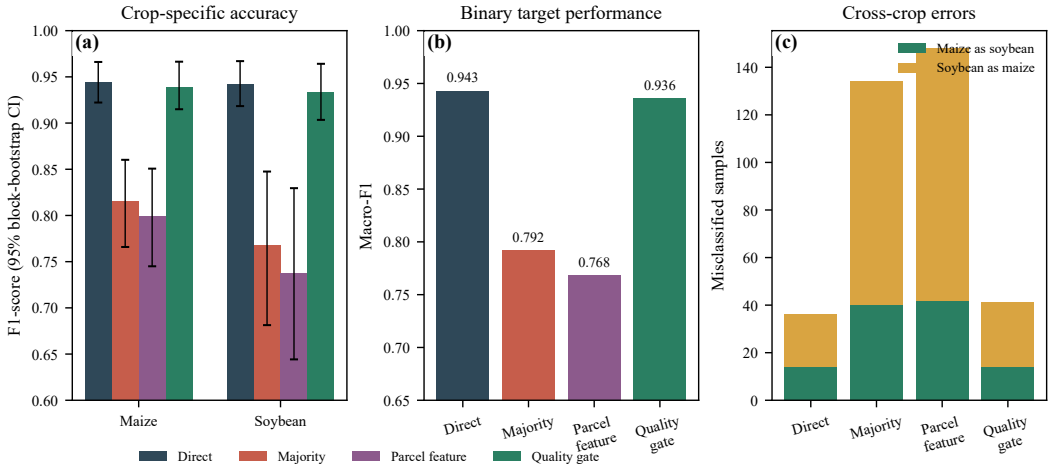

Figure 7. Maize–soybean accuracy under the four attribute strategies: (**a**) crop-specific F1 with 95% spatial-block bootstrap intervals, (**b**) macro-F1, and (**c**) cross-crop error counts.

Table 5. Maize–soybean classification accuracy. F1 intervals are 95% spatial-block bootstrap confidence intervals.

Method	Crop	Precision	Recall	F1 [95% CI]	Macro-F1
Direct AEF	Maize	0.936	0.952	0.944 [0.922, 0.966]	0.943
Direct AEF	Soybean	0.955	0.930	0.942 [0.918, 0.967]	0.943
Majority homogenization	Maize	0.759	0.881	0.815 [0.766, 0.860]	0.792
Majority homogenization	Soybean	0.847	0.703	0.768 [0.681, 0.848]	0.792
Parcel-feature clustering	Maize	0.735	0.875	0.799 [0.745, 0.851]	0.768
Parcel-feature clustering	Soybean	0.833	0.661	0.737 [0.644, 0.830]	0.768
Quality-gated refinement	Maize	0.922	0.955	0.939 [0.915, 0.966]	0.936
Quality-gated refinement	Soybean	0.954	0.915	0.934 [0.903, 0.964]	0.936

Relative to direct AEF, the macro-F1 difference was −0.151 for majority homogenization, with a 95% confidence interval of −0.197 to −0.105, and −0.175 for parcel-feature clustering, with an interval of −0.227 to −0.128. The corresponding exact McNemar p values were 2.16×10^{-21} and 2.31×10^{-25}. For quality-gated refinement, the macro-F1 difference was −0.007, with an interval of −0.017 to 0.002 and a McNemar p value of 0.219.

Compared with majority homogenization, quality-gated refinement had a macro-F1 difference of 0.144, with a 95% confidence interval of 0.103–0.196. Negative block-level differences for the unconditional strategies occurred across multiple validation blocks, whereas most quality-gated differences were close to zero. Figure 8 and Table 6 show the paired comparisons.

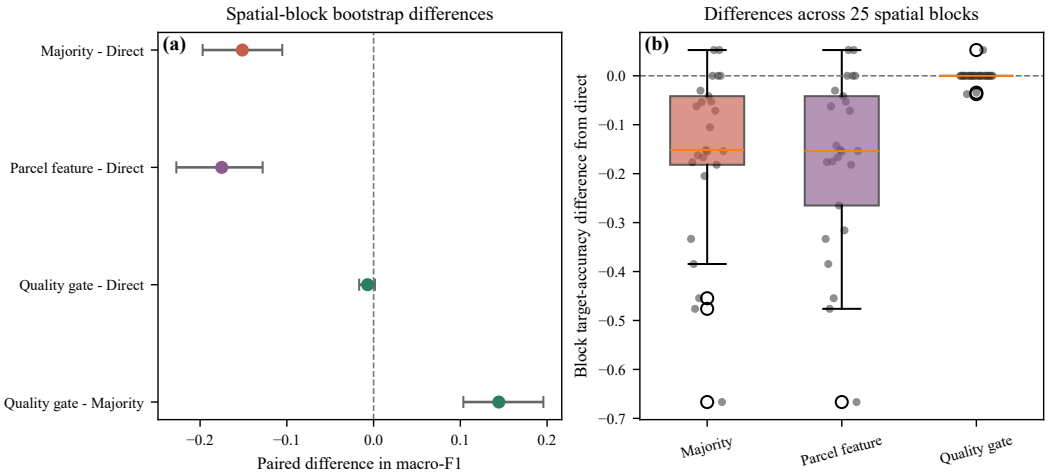

Figure 8. Paired comparisons among mapping strategies. (**a**) Spatial-block bootstrap differences in macro-F1 and (**b**) target-accuracy differences from direct AEF across the 25 validation blocks.

Table 6. Paired differences between mapping strategies.

Comparison	ΔMacro-F1 [95% CI]	ΔMaize F1	ΔSoybean F1	McNemar p
Majority − Direct	−0.151 [−0.197, −0.105]	−0.129	−0.174	2.16×10^{-21}
Parcel feature − Direct	−0.175 [−0.227, −0.128]	−0.145	−0.205	2.31×10^{-25}
Quality gate − Direct	−0.007 [−0.017, +0.002]	−0.005	−0.009	0.219
Quality gate − Majority	+0.144 [+0.103, +0.196]	+0.123	+0.166	7.84×10^{-21}

At the common reference locations, majority homogenization converted 106 predictions that were correct under direct AEF into errors. These comprised 75 soybean observations changed to maize and 31 maize observations changed to soybean. The correct-to-wrong rate was 38.0% in fields with direct purity below 0.75, 19.7% at purity 0.75–0.90, and zero above

0.95. Fields of at least 50 ha contributed 63.2% of the new errors, and fields of at least 10 ha contributed 92.5%. Figure 9 relates the new errors to field purity, area, and crop transition.

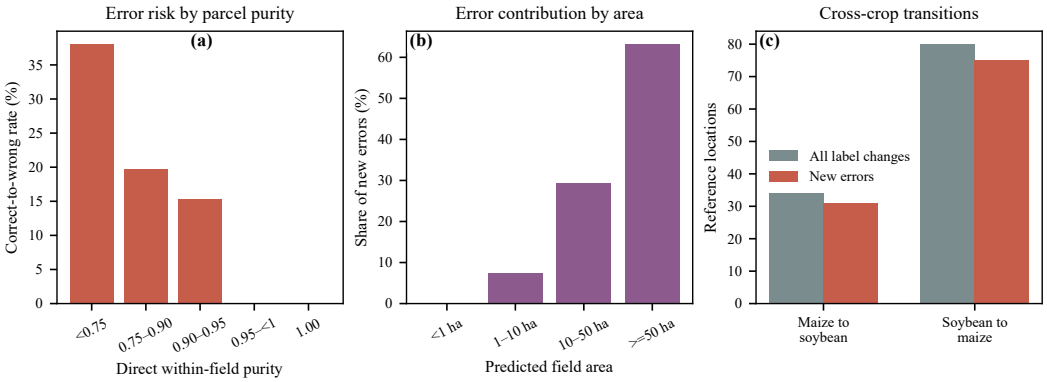

Figure 9. Errors introduced by majority homogenization at the reference locations. (**a**) Proportion of observations correctly classified by direct AEF that became errors, grouped by within-field purity; (**b**) contribution of field-area groups to the new errors; and (**c**) maize–soybean label transitions. The diagnostics are not area-adjusted regional estimates.

3.6. Sensitivity of the Quality-Gating Strategy

Across 80 combinations of purity, minimum pixel support, and heterogeneity filtering, accepted-field fractions ranged from 36.7% to 84.6%, changed-pixel fractions from 0.057% to 1.911%, and macro-F1 from 0.921 to 0.943. At $n_{\min} = 10$ without an explicit heterogeneity threshold, increasing the purity threshold from 0.80 to 0.95 reduced the accepted-field fraction from 69.2% to 57.5% and the changed-pixel fraction from 1.911% to 0.303%. Macro-F1 increased from 0.921 to 0.943. Varying minimum support from 3 to 20 changed the accepted-field count but had little effect on the changed-pixel fraction or macro-F1.

The primary operating point ($\tau = 0.90$, $n_{\min} = 10$, no additional heterogeneity threshold) accepted 62.7% of fields, changed 0.704% of pixels, and yielded a macro-F1 of 0.936. The stricter $\tau = 0.95$ setting accepted 57.5% of fields and changed 0.303% of pixels, with a macro-F1 of 0.943. Thus, the primary setting allowed more parcel-level modification while yielding slightly lower agreement at the sampled locations.

At the primary purity and support thresholds, applying the 75th-percentile heterogeneity threshold reduced the accepted-field fraction to 49.0% and the changed-pixel fraction to 0.090%, while macro-F1 equalled that of direct AEF. Without an explicit heterogeneity threshold, macro-F1 was 0.936 and a larger pixel fraction was modified. Standardized 64-dimensional features and PCA representations retaining 90%, 95%, and 99% of variance all had ARI values of at least 0.9991 relative to the primary parcel partition and produced the same sample macro-F1. Figure 10 and Table 7 summarize these sensitivity results.

Table 7. Selected quality-gate sensitivity results.

Purity	Min. Pixels	Heterogeneity	Accepted Fields (%)	Changed Pixels (%)	Macro-F1
0.80	10	none	69.2	1.911	0.921
0.85	10	none	66.2	1.204	0.935
0.90	10	none	62.7	0.704	0.936
0.90	10	q75	49.0	0.090	0.943
0.90	10	q90	60.2	0.542	0.936
0.95	10	none	57.5	0.303	0.943

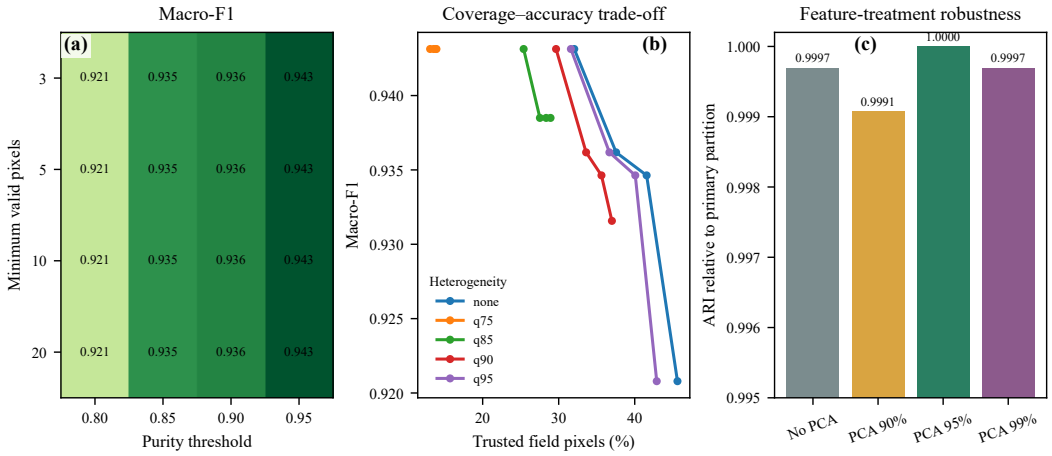

Figure 10. Sensitivity analyses: (**a**) macro-F1 across purity and minimum-support thresholds, with darker green indicating higher macro-F1, (**b**) coverage–accuracy trade-off under alternative heterogeneity thresholds (none: blue; q75: orange; q85: green; q90: red; q95: purple), and (**c**) stability under alternative AEF feature treatments.

In the gate ablation, the leave-one-component variants separated the effects of purity, minimum support, and the under-segmentation screen. Removing any one condition increased the accepted-field fraction relative to the complete gate. Figure 11 compares their effects on spatial modification and error control.

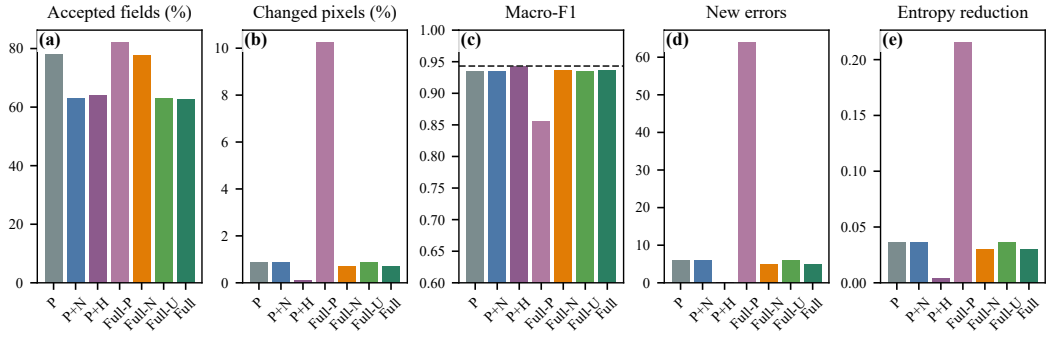

Figure 11. Quality-gate ablation. P denotes purity ≥ 0.90; P + N adds at least 10 valid pixels; P + H adds a 75th-percentile feature-heterogeneity screen. Full-P, Full-N, and Full-U are leave-one-component variants of the complete gate with purity, minimum support, and under-segmentation exclusion removed, respectively; Full retains all three conditions. Panels report (**a**) accepted-field fraction, (**b**) modified-pixel fraction, (**c**) macro-F1, (**d**) observations correct under direct AEF and incorrect after refinement, and (**e**) reduction in pixel-weighted field entropy. The dashed line in panel (**c**) marks direct AEF.

3.7. Sensitivity to Spatial-Block Size

The 0.025°, 0.05°, and 0.10° grids produced 52, 25, and 11 occupied blocks, respectively (Table A2). All five strategies retained the same crop naming in every training fold at all three scales, with no ties, and their out-of-fold predictions were unchanged. Direct AEF and quality-gated macro-F1 therefore remained 0.943 and 0.936. Their paired difference was −0.007, with 95% intervals of [−0.019, 0.002], [−0.017, 0.002], and [−0.015, 0.003] from finer to coarser grids. Although point estimates were stable, interval limits and fold balance varied. The coarsest grid had only 11 blocks and 94–210 observations per held-out fold, warranting caution in interpreting its bootstrap uncertainty.

4. Discussion

4.1. Discriminative Capacity of AEF Embeddings for Maize and Soybean

The direct-pixel results indicate that annual AEF embeddings retained crop-relevant structure before crop labels were introduced. Two pixel clusters were associated with maize and soybean, while repeated fits and spatial subsamples showed a stable parcel-feature partition (Figure 4). This pattern accords with the design of AEF as a reusable representation of multi-source, multi-temporal Earth-observation data [14,15]. Agricultural benchmarks and supervised embedding-based crop mapping have examined other downstream uses [16,17]. Here, clustering provides the feature partition and reference observations supply crop names; differences in crops, regions, sampling, and adaptation methods limit direct numerical comparisons with those studies.

Differences in crop development, canopy structure, planting and maturity dates, and field management may contribute to the separation observed in the annual embeddings. Related temporal information is central to crop discrimination in conventional satellite time series [2–4]. AEF encodes spatial and temporal relationships across multiple observation sources [14,15], but the annual product used here supplies one 64-dimensional vector per pixel, not a dated spectral or vegetation-index sequence. Its dimensions cannot directly identify the contribution of individual processes or provide an explicit phenological curve.

Transfer across sowing dates, rotations, multiple-cropping systems, climates, and management practices depends on which seasonal cues drive these embedding–crop relationships. The present analysis does not resolve those cues: phenological reconstruction, crop-stage attribution, and controlled comparisons with seasonal Sentinel-2 time series were outside its scope. Tests across crop calendars and management systems would help determine when an annual embedding retains sufficient temporal detail for crop discrimination.

Parcel entropy offers a narrower diagnostic: it summarizes the consistency of cluster labels within a predicted field. Entropy has also been used to characterize high-dimensional biological data, including pathway activation from gene-expression profiles [37]. In this application, it measures within-field label diversity without identifying the crop processes encoded by AEF.

The four clusters represented patterns in the embedding space rather than four crop classes. The $K = 5$ analysis in Figure 12 changed the detailed partition but retained similar maize–soybean associations and rejection behavior. This result is compatible with a many-to-one relation between feature clusters and crop classes, in which a crop may occupy more than one part of the embedding space and poorly supported clusters remain unassigned. Reference observations give these feature groups their crop meaning.

4.2. Error Propagation Under Unconditional Parcel Processing

Figures 6 and 7 show that both unconditional parcel strategies produced more spatially regular maps but lower maize–soybean accuracy at the reference locations. Majority homogenization removed small fragments by assigning one existing pixel cluster to each field. Parcel-feature clustering also changed the feature distribution by fitting a new model to parcel means. The visual regularity of these products was therefore not accompanied by better thematic agreement with the reference observations.

The reference-point audit helps characterize this loss. Figure 9 shows that new errors under majority homogenization occurred more often from soybean to maize than in the opposite direction and were concentrated in large, low-purity FTW polygons. This pattern is consistent with a dominant label replacing minority pixel patterns within mixed polygons and may account for part of the decline in soybean recall.

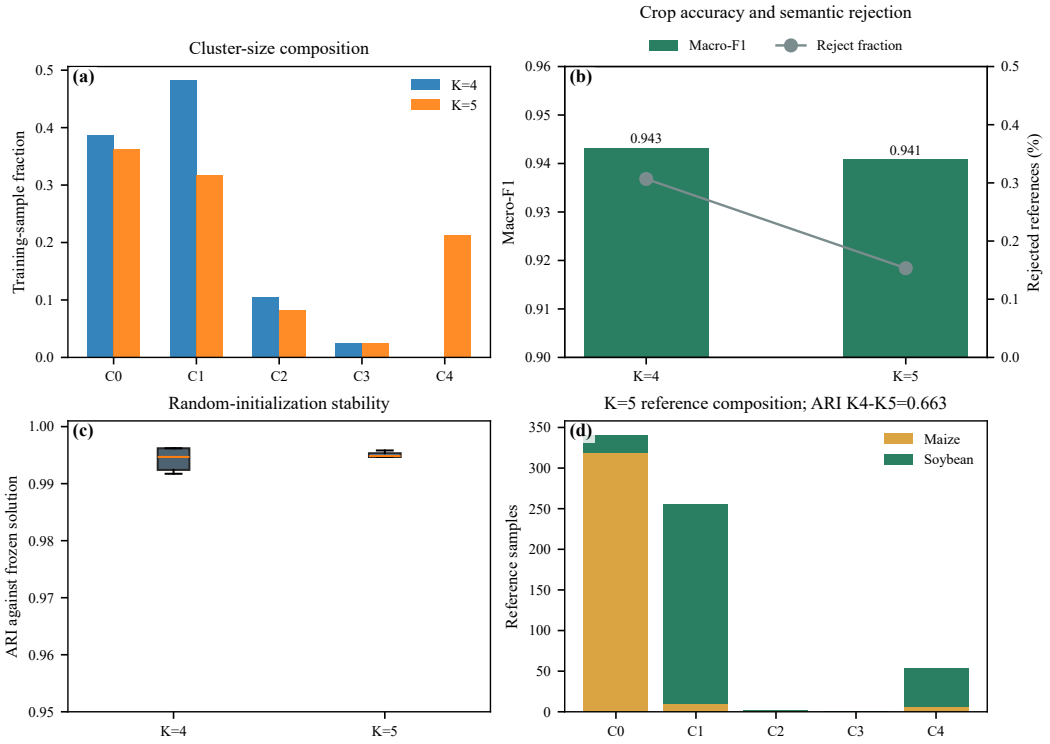

Figure 12. Sensitivity of the direct AEF structure to cluster count. (**a**) Cluster fractions in the common 250,000-pixel sample; (**b**) out-of-fold maize–soybean macro-F1 and rejected-reference fraction; (**c**) agreement with the fixed solution over 12 repeated fits; and (**d**) maize and soybean reference composition of the $K = 5$ clusters. Crop labels were not used to fit either clustering model.

Figure 13 provides complementary spatial evidence from four manually inspected parcels. Each was a large, low-purity footprint flagged for possible under-segmentation. The polygons crossed several coherent AEF patterns that remained visible in the direct map but were replaced by one label under unconditional parcel processing. None of these parcels passed the quality gate. The cases illustrate a plausible error pathway, but they are not an independent boundary-validation sample. They were selected from observed errors, the AEF composite is not optical true color, and authoritative local boundaries were unavailable.

Together, these observations suggest that some predicted footprints did not correspond to single crop-management units. A footprint may include adjacent fields, crop strips, roads, or shelterbelts, allowing a majority operation to extend a local classification error across the polygon. FTW provides a global benchmark and broad access to predicted field geometry [20,21], but individual predictions may still contain segmentation errors in a given landscape. Many object-based crop studies instead use authoritative declarations, cadastral units, or locally optimized image objects [18,19,38]. The suitability of parcel-level mapping therefore depends partly on how closely the analysis object corresponds to a crop-management unit.

Parcel-feature clustering introduced an additional change by shifting the unit of analysis from pixels to area-weighted parcel means. Averaging may suppress local variation that contributes to crop separation, especially when a footprint contains several coherent subregions. Its difference from direct clustering consequently reflects both parcel-wide assignment and a new partition of the aggregated feature space.

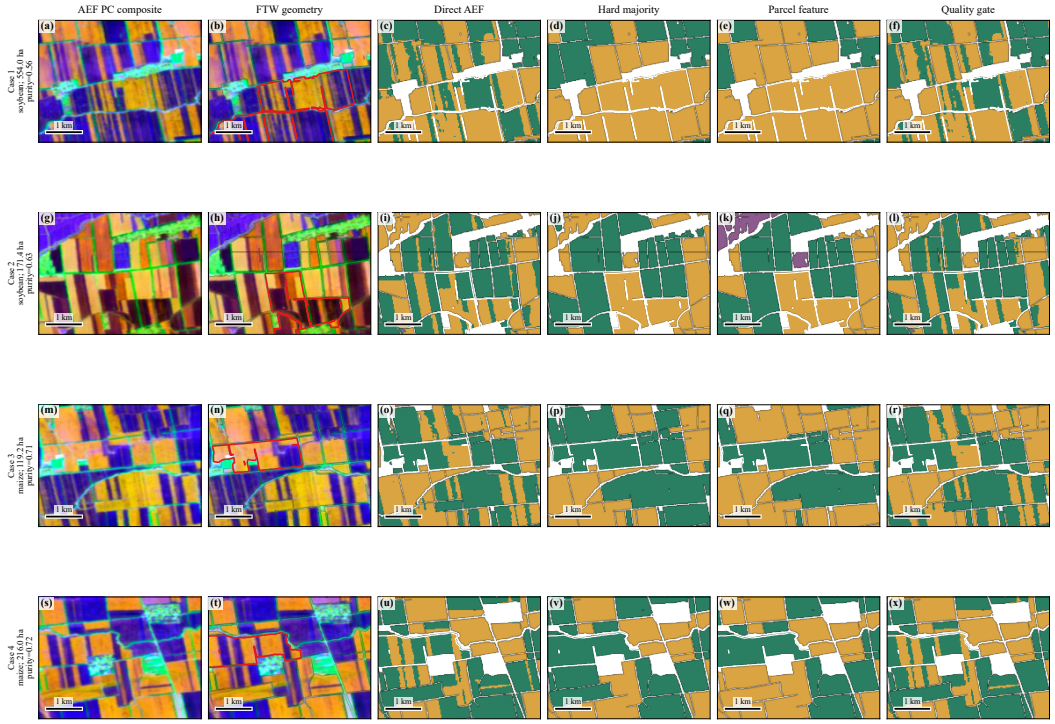

Figure 13. Representative low-purity fields associated with new errors after unconditional parcel processing. Rows (**a**–**f**), (**g**–**l**), (**m**–**r**), and (**s**–**x**) show Cases 1–4, respectively. Within each row, the six columns show an AEF three-component embedding composite, the composite with FTW geometry, direct AEF clusters, majority homogenization, parcel-feature clustering, and quality-gated refinement. Cluster-map colors denote cluster 0 (ochre), cluster 1 (green), cluster 2 (gray-blue), and cluster 3 (purple); colors in the embedding composite encode AEF feature values. White map areas lie outside the displayed FTW support. The composite is not optical true color. Red lines identify the focal FTW footprint, and white lines show field boundaries in the cluster maps. Rows contain two soybean and two maize cases that were correct under direct AEF and incorrect after unconditional parcel processing.

4.3. Geometric Masking and Quality-Gated Parcel Refinement

The comparison separates two uses of predicted polygons. The full-domain Direct AEF map provides the reference for measuring the loss of spatial support, while the common-domain Direct AEF view provides the reference for within-footprint attribute comparisons. Geometric masking changed the output domain while preserving retained pixel labels, whereas hard homogenization changed attributes within that domain. A boundary may therefore help delimit the mapped agricultural extent without providing sufficient evidence for a single crop label. This difference between geometric and attribute use is important when predicted rather than authoritative fields are available.

The quality gate applied the direct-pixel majority only when pixel support, direct-label purity, and the under-segmentation screen met the specified criteria. It changed a small fraction of pixels, reduced within-field fragmentation, and remained close to the direct AEF result. The error audit and inspected cases indicate that the gate excluded several low-purity or potentially under-segmented polygons in which majority assignment introduced errors. In the tested setting, the gate limited parcel-induced error propagation by preserving pixel evidence in polygons unsuitable for uniform assignment.

The ablation results in Figure 11 indicate that purity alone did not reproduce the behavior of the primary gate. Purity-only variants changed more pixels and produced more correct-to-wrong transitions, whereas the primary gate limited these transitions while reducing parcel entropy. In the primary gate, feature heterogeneity entered through

the under-segmentation screen rather than through a global threshold. It was considered together with support, purity, and morphology. Figure 14 shows that accepted parcels could therefore have higher marginal heterogeneity than rejected parcels; selection depended on the combination of criteria rather than on feature variance alone.

Figure 10 and Table 7 place the primary gate within the observed trade-off between parcel-level modification and agreement with the reference observations. Stricter settings retained more direct-pixel labels and achieved higher sample macro-F1 while allowing less parcel-wide refinement. The primary thresholds provide an empirical operating point for this balance. Field size, crop mixture, and boundary quality can guide the choice of settings in other landscapes.

The 31.0% reduction in small patches describes a net change in eight-neighbor label components smaller than ten pixels, rather than a count of verified spurious fragments or physical fields removed. Components can connect across predicted polygon boundaries, and minority crop patterns may be overwritten even in accepted parcels. The descriptive area-stratified audit in Table A3 found no new gate-induced sample errors below 10 ha, but only five reference observations were associated with polygons smaller than 1 ha. This limited sample cannot establish preservation of real small fields, and the available reference data cannot determine whether the reduction predominantly reflects removal of errors.

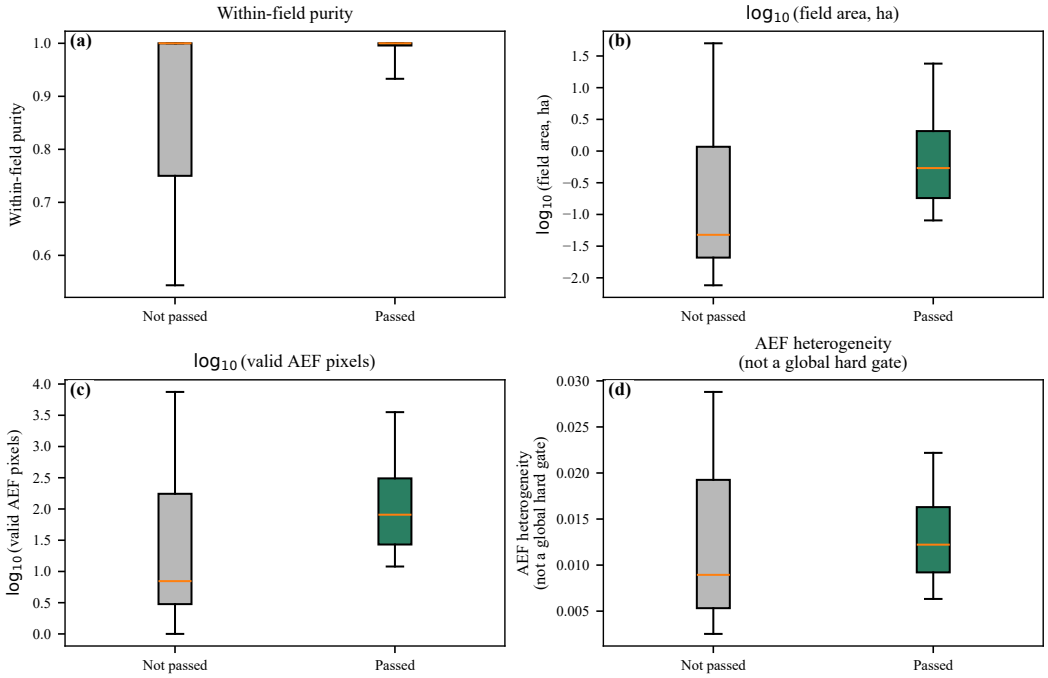

Figure 14. Characteristics of fields accepted and rejected by the primary gate: (**a**) direct-cluster purity, (**b**) field area, (**c**) valid AEF pixel support, and (**d**) mean AEF feature heterogeneity. Heterogeneity was used within the under-segmentation screen rather than as a global threshold, so accepted fields can have a higher marginal heterogeneity distribution. Boxes show interquartile ranges, center lines show medians, and whiskers show the 5th and 95th percentiles.

4.4. Relationship to Existing Crop-Mapping Approaches and Applications

Pixel- and parcel-level maps retain different aspects of the landscape. Pixel models preserve local transitions and possible boundary mismatches but often contain mixed edge pixels and small fragments. Parcel models provide compact units for management applications but rely on boundaries that approximate internally homogeneous objects. FTW expands access to field geometry where cadastral data are unavailable [20,21]. The

present results indicate that uncertainty in this geometry remains relevant when it is used to constrain crop attributes.

The quality-gated approach combines pixel and parcel information rather than selecting one spatial unit for the entire map. AEF pixels provide the category evidence, while parcel diagnostics determine where a uniform label is applied. The output can distinguish uniformly labeled parcels, parcels that retain pixel predictions, and areas with rejected semantic assignments. Gate status, purity, pixel support, and under-segmentation flags may also assist manual review and subsequent uncertainty analysis.

Unconditional majority-vote homogenization supplies an object-based post-processing baseline on the same FTW geometry. Its comparison with quality gating isolates the effect of selective replacement. Extending this comparison to moving-window majority and connected-component filters would help assess whether parcel geometry adds value over local raster smoothing.

The same integration principle could be examined with other representations, classifiers, or boundary products. Supervised class probabilities could replace direct clusters, and regional segmentation could replace FTW. Uncertainty-aware crop classifiers and conditional random fields provide ways to retain probabilistic evidence while encouraging spatial agreement [24,25]. Continuous parcel weights or graph-based constraints could similarly allow field information to influence pixel predictions more gradually than the binary gate.

For a parcel graph, fields could form nodes linked by adjacency or feature similarity. Cross-domain examples include graph learning for herb–disease associations [39] and network enhancement for disease-gene identification [40], both of which integrate information across heterogeneous networks. These examples suggest possible ways to combine field geometry and embedding similarity; their usefulness for crop mapping would require a separate evaluation.

4.5. Limitations and Future Research

The analysis covers one landscape window and one year. Additional years and regions are needed to assess whether cluster semantics, feature separability, and gate thresholds transfer across climates, field-size distributions, and management systems. The reference data contain only maize and soybean, so rejected clusters cannot be assigned to forest, settlement, rice, or other classes. The reported metrics describe maize–soybean retrieval at the reference locations rather than all-class land-cover accuracy.

Sparse target-crop support cannot distinguish other land covers from mixed pixels or atypical maize–soybean conditions. A semantically rejected location therefore indicates unresolved crop identity, not confirmed absence of maize or soybean. This differs from geometric masking outside FTW footprints: within the mask, final cluster labels 2 and 3 remain outside the named two-crop output. Failing the quality gate does not remove a parcel; its direct-pixel labels are retained. The resulting crop layer therefore need not cover every pixel within the retained FTW domain, and low rejection at the sampled crop points does not establish spatial completeness across that domain.

Reference observations were concentrated near survey routes. Spatial folding and block bootstrap reduced local dependence but did not make this convenience sample representative of the mapped region. Point-based F1-scores and error shares among sampled parcels do not support crop-area estimation or regional estimates of parcel defects. Probability sampling and area-adjusted designs remain necessary for unbiased map-accuracy and crop-area inference [34,35]. Spatial separation can provide a more demanding transfer assessment, but it does not resolve sample-selection bias [11,41].

No independent local field-boundary reference was available. The inspected parcels and under-segmentation diagnostics identify plausible boundary mismatches but cannot distinguish boundary displacement from mixed management or embedding misclassification, nor estimate their prevalence. Comparison with authoritative or systematically interpreted boundaries would allow direct estimation of under-segmentation and over-segmentation rates and analysis of thematic accuracy near predicted edges.

Parcel-wide replacement remains a coarse operation even when gated. A continuous confidence measure combining boundary confidence, pixel-class uncertainty, embedding heterogeneity, neighborhood context, and polygon morphology could offer finer control over where labels change.

5. Conclusions

In the tested single-year maize–soybean landscape, annual AlphaEarth Foundations embeddings separated crop-relevant feature groups without crop labels entering dimensionality reduction or clustering. Following foldwise semantic assignment, direct AEF clustering achieved F1-scores of 0.944 for maize and 0.942 for soybean at the 652 route-based reference locations.

Fields of the World boundaries supplied spatial units rather than the crop information itself. Unconditional homogenization produced regular field patterns but propagated errors where predicted polygons contained several AEF feature groups. The proposed quality gate used pixel support, within-field purity, and under-segmentation risk to identify polygons suitable for parcel-wide assignment. At the same route-based reference locations, selective integration yielded a macro-F1 of 0.936, compared with 0.943 for direct AEF clustering, while reducing within-field fragmentation in the mapped domain.

Annual embeddings and globally predicted boundaries allow this workflow to be applied without reconstructing local image time series or obtaining cadastral data, while reference labels are still needed for crop naming and assessment. This provides a basis for testing larger-area applications. Evaluation across regions and years can guide the adaptation of gate thresholds to different crop systems and field configurations.

Author Contributions: Conceptualization, R.K. and J.L.; methodology, R.K.; software, R.K.; validation, R.K.; formal analysis, R.K.; investigation, R.K.; resources, R.K. and R.B.; data curation, R.K.; writing—original draft preparation, R.K.; writing—review and editing, R.B. and J.L.; visualization, R.K.; supervision, J.L.; project administration, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the New Generation Artificial Intelligence-National Science and Technology Major Project of China (grant number 2026ZD0125100).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The annual AEF product is available through the Google Earth Engine Data Catalog. Access to FTW is described by Kerner et al. [20]. Derived products and analysis code will be deposited in a persistent public archive. Access to the field observations is subject to the data provider's terms.

Acknowledgments: The authors thank the providers of the AlphaEarth Foundations product, the Fields of the World dataset, and the 2024 field survey data used in this study.

Conflicts of Interest: Authors Runze Kang and Ran Bi are employed by the company China Telecom Artificial Intelligence Technology (Beijing) Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Supplementary Methods

Appendix A.1. Data Parsing and Computational Implementation

The field-survey table was stored as a GB18030-encoded CSV with 29 nominal columns. In 1600 records, unquoted commas in the final structured field produced more than 28 delimiters. These rows were repaired by retaining the first 28 delimiters as the fixed schema and joining all remaining fragments into the final field before parsing. The repaired table was then subjected to the year, coordinate, crop-name, and study-coverage quality controls described in Section 2.2.3.

The 15 AEF GeoTIFF tiles were combined without resampling in a virtual raster. The native mosaic used EPSG:4326, contained 19,237 columns by 11,469 rows, and stored NoData as −9999. Full-scene prediction was streamed in 512 × 512 raster blocks with four worker threads; these settings affected memory use and run time but not model definition or numerical results.

Random seeds were fixed at 20240724 for pixel sampling and 20240729 for cluster-count sensitivity. The saved accuracy analysis used seeds 20240728–20240732 for single-product block-bootstrap intervals in the order Direct AEF, geometry masking, majority homogenization, parcel-feature clustering, and quality gating, and 20240828–20240831 for the four paired comparisons in their reported order. Parcel stability fits used deterministic offsets from the primary sequence. The retained parcel solution used seed 20240775 and combined a low K-means objective with a mean pairwise ARI of 0.9884.

One reference point intersected the rasterized FTW mask under the pixel-center rule but did not intersect the source vector under exact point-in-polygon matching. It remained in the 652-point common raster comparison and was excluded only from sensitivity analyses requiring vector parcel attributes, which therefore used 651 points.

For cartographic display, a boundary-cut version removed a one-cell line along shared polygon boundaries so that neighboring fields assigned the same crop remained visible as separate objects. The resulting NoData boundary cells were excluded from pixel counts, area summaries, accuracy calculations, and structural comparisons.

Appendix A.2. Foldwise Evidence for Crop Semantic Assignment

Table A1 lists the training-fold counts used to name the fixed direct-pixel clusters. Dominant-crop proportions summarize the composition of the training observations; the primary $K = 4$ mapping used the fixed candidate rule in Section 2.6.

Table A1. Training-fold evidence for the semantic assignment of the fixed direct-pixel $K = 4$ partition. Fold identifiers are zero-based. Train/test gives the numbers of training and held-out observations. Dominant (%) is the larger crop count divided by the cluster's training support; a dash indicates no support. Clusters 2 and 3 were rejected by the fixed candidate-set rule, not by a fold-specific dominance threshold.

Fold	Train/Test	Cluster	Maize	Soybean	Support	Dominant (%)	Assigned
0	521/131	0	265	16	281	94.31	Maize
		1	11	228	239	95.40	Soybean
		2	1	0	1	100.00	Reject
		3	0	0	0	–	Reject
1	523/129	0	257	15	272	94.49	Maize
		1	9	240	249	96.39	Soybean
		2	2	0	2	100.00	Reject
		3	0	0	0	–	Reject
2	519/133	0	248	20	268	92.54	Maize
		1	13	237	250	94.80	Soybean
		2	1	0	1	100.00	Reject
		3	0	0	0	–	Reject

Table A1. *Cont*.

Fold	Train/Test	Cluster	Maize	Soybean	Support	Dominant (%)	Assigned
3	522/130	0	240	19	259	92.66	Maize
		1	12	249	261	95.40	Soybean
		2	2	0	2	100.00	Reject
		3	0	0	0	–	Reject
4	523/129	0	270	18	288	93.75	Maize
		1	11	222	233	95.28	Soybean
		2	2	0	2	100.00	Reject
		3	0	0	0	–	Reject

Appendix A.3. Spatial-Block Size Sensitivity

At each grid scale, blocks were sampled with replacement, retaining all observations within each selected block; the 2.5th and 97.5th percentiles defined the intervals. Bootstrap replicates used fixed out-of-fold predictions without refitting crop naming. Per-strategy random seeds were retained across scales, so identical predictions can have slightly different Monte Carlo interval limits. The paired direct-versus-gated analysis used seed 20240830 and shared block draws. The coarser grid contains fewer resampling units; narrower intervals at that scale should therefore be interpreted alongside block counts and fold balance. All scales share the route-based sample and lack buffers between neighboring blocks.

Table A2. Spatial-block size sensitivity with a fixed grid origin and 652 reference observations. Strategy rows report macro-F1 with 95% block-bootstrap intervals in brackets; the final row reports the paired macro-F1 difference. All scales use five folds and 2000 bootstrap replicates. The 0.05° column reproduces the primary analysis.

Quantity or Strategy	0.025°	0.05° (Primary)	0.10°
Occupied blocks	52	25	11
Observations per block	1–50	1–83	3–210
Observations per test fold	130–131	129–133	94–210
Direct AEF	0.943 [0.922, 0.964]	0.943 [0.924, 0.965]	0.943 [0.928, 0.961]
Geometry mask	0.943 [0.923, 0.964]	0.943 [0.923, 0.966]	0.943 [0.928, 0.961]
Majority homogenization	0.792 [0.732, 0.845]	0.792 [0.736, 0.845]	0.792 [0.726, 0.867]
Parcel-feature clustering	0.768 [0.705, 0.824]	0.768 [0.708, 0.826]	0.768 [0.710, 0.835]
Quality-gated refinement	0.936 [0.911, 0.961]	0.936 [0.913, 0.964]	0.936 [0.917, 0.959]
Gated minus direct	−0.007 [−0.019, 0.002]	−0.007 [−0.017, 0.002]	−0.007 [−0.015, 0.003]

Appendix A.4. Descriptive Audit by Predicted Polygon Area

The 652 out-of-fold predictions were grouped by the area of their raster-assigned FTW polygon into <1, 1–<5, 5–<10, 10–<50, and ≥50 ha classes. The observations occupied 211 distinct predicted polygons. The point without exact vector containment described in Appendix A.1 was assigned to its raster polygon's area group. Table A3 reports the fraction classified correctly, counting rejected predictions as incorrect, and correct-to-wrong transitions relative to Direct AEF.

The gate introduced no new sample errors among the 138 observations assigned to polygons smaller than 10 ha and introduced two and three in the 10–<50 and ≥50 ha groups, respectively. The <1 ha group contained only five observations (four maize and one soybean); all three strategies classified four correctly. These descriptive proportions weight observations rather than fields, and multiple observations can share a polygon. They neither estimate representative accuracy by field size nor identify the true status of changed raster components. Polygon area is a predicted-object attribute, not an independent measurement of actual management-field size.

Table A3. Descriptive sample performance by raster-assigned FTW polygon area. D, M, and G denote Direct AEF, majority homogenization, and quality-gated refinement. Correct (%) uses all observations in the group as the denominator, including rejected predictions. C → W counts observations correct under D and incorrect under the indicated strategy. Polygons counts distinct sampled predicted polygons; the rows do not represent independent field-validation samples.

Area (ha)	Samples	Polygons	D Correct (%)	M Correct (%)	G Correct (%)	M C → W	G C → W
<1	5	5	80.0	80.0	80.0	0	0
1–<5	69	36	95.7	95.7	97.1	2	0
5–<10	64	32	93.8	84.4	93.8	6	0
10–<50	275	94	94.9	85.5	94.2	31	2
≥50	239	44	93.3	66.5	92.1	67	3
All	652	211	94.2	79.4	93.6	106	5

References

1. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
2. Wardlow, B.D.; Egbert, S.L. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* **2008**, *112*, 1096–1116. [CrossRef]
3. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]
4. Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.-F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [CrossRef]
5. You, N.; Dong, J.; Huang, J.; Du, G.; Zhang, G.; He, Y.; Yang, T.; Di, Y.; Xiao, X. The 10-m crop type maps in Northeast China during 2017–2019. *Sci. Data* **2021**, *8*, 41. [CrossRef] [PubMed]
6. Liu, Y.; Wang, J. Revealing Annual Crop Type Distribution and Spatiotemporal Changes in Northeast China Based on Google Earth Engine. *Remote Sens.* **2022**, *14*, 4056. [CrossRef]
7. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
8. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
9. Pelletier, C.; Webb, G.; Petitjean, F. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]
10. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
11. Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecol. Model.* **2019**, *411*, 108815. [CrossRef]
12. Reed, C.J.; Gupta, R.; Li, S.; Brockman, S.; Funk, C.; Clipp, B.; Keutzer, K.; Candido, S.; Uyttendaele, M.; Darrell, T. Scale-MAE: A Scale-Aware Masked Autoencoder for Multiscale Geospatial Representation Learning. In *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 4065–4076. [CrossRef]
13. Mendieta, M.; Han, B.; Shi, X.; Zhu, Y.; Chen, C. Towards Geospatial Foundation Models via Continual Pretraining. In *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 16760–16770. [CrossRef]
14. Brown, C.F.; Kazmierski, M.R.; Pasquarella, V.J.; Rucklidge, W.J.; Samsikova, M.; Zhang, C.; Shelhamer, E.; Lahera, E.; Wiles, O.; Ilyushchenko, S.; et al. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv* **2025**, arXiv:2507.22291.
15. Google Earth Engine Data Catalog. Google Satellite Embedding V1 Annual. 2025. Available online: https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_SATELLITE_EMBEDDING_V1_ANNUAL (accessed on 28 July 2026).
16. Ma, Y.; Shen, Y.; Swatantran, A.; Lobell, D.B. Harvesting AlphaEarth: Benchmarking the geospatial foundation model for agricultural downstream tasks. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *149*, 105258. [CrossRef]

17. Narimani, M.; Pourreza, A.; Farajpoor, P. Mapping Tomato Cropping Systems in California Using AlphaEarth Geospatial Embeddings and Deep Learning Analysis. In *Proceedings of the 2026 ASABE Annual International Meeting, Indianapolis, Indiana, 12–15 July 2026*; American Society of Agricultural and Biological Engineers: St. Joseph, MI, USA, 2026. [CrossRef]
18. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
19. Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [CrossRef]
20. Kerner, H.; Chaudhari, S.; Ghosh, A.; Robinson, C.; Ahmad, A.; Choi, E.; Jacobs, N.; Holmes, C.; Mohr, M.; Dodhia, R.; et al. Fields of The World: A Machine Learning Benchmark Dataset for Global Agricultural Field Boundary Segmentation. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 28151–28159. [CrossRef]
21. Fields of the World. Global Field Boundary Predictions: China, Heilongjiang (cn_hl), Vector Alpha Release. Source Cooperative. Available online: https://data.source.coop/ftw/global-data/predictions/vectors/alpha/results-by-admin-conf/admin:country_code=CN/CN_HL.json (accessed on 28 July 2026).
22. Persello, C.; Tolpekin, V.A.; Bergado, J.R.; de By, R.A. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* **2019**, *231*, 111253. [CrossRef] [PubMed]
23. Waldner, F.; Diakogiannis, F.I. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.* **2020**, *245*, 111741. [CrossRef]
24. Lu, Q.; Zhao, W.; Chen, J.; Chen, X.; Zhang, L. Uncertainty Mixture of Experts Model for Long Tail Crop Type Mapping. *Remote Sens.* **2025**, *17*, 3752. [CrossRef]
25. Zhu, Q.; Jiang, Y.; Fan, L. Classwise-CRF: Category-specific fusion for enhanced semantic segmentation of remote sensing imagery. *Neural Netw.* **2026**, *197*, 108485. [CrossRef] [PubMed]
26. Li, R.; Wang, Y.; Sun, S.; Zhang, Y.; Ding, F.; Gao, H. UE-Extractor: A Grid-to-Point Ground Extraction Framework for Unstructured Environments Using Adaptive Grid Projection. *IEEE Robot. Autom. Lett.* **2025**, *10*, 5991–5998. [CrossRef]
27. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]
28. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
29. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
30. Calinski, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Commun. Stat. Simul. Comput.* **1974**, *3*, 1–27. [CrossRef]
31. Davies, D.L.; Bouldin, D.W. A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. [CrossRef]
32. Fraley, C.; Raftery, A.E. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.* **2002**, *97*, 611–631. [CrossRef]
33. Hubert, L.; Arabie, P. Comparing partitions. *J. Classif.* **1985**, *2*, 193–218. [CrossRef]
34. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [CrossRef]
35. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
36. McNemar, Q. Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika* **1947**, *12*, 153–157. [CrossRef] [PubMed]
37. Li, X.; Li, M.; Xiang, J.; Zhao, Z.; Shang, X. SEPA: Signaling entropy-based algorithm to evaluate personalized pathway activation for survival analysis on pan-cancer data. *Bioinformatics* **2022**, *38*, 2536–2543. [CrossRef] [PubMed]
38. Defourny, P.; Bontemps, S.; Bellemans, N.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Nicola, L.; Rabaute, T.; et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* **2019**, *221*, 551–568. [CrossRef]
39. Hu, X.; Lu, Y.; Tian, G.; Bing, P.; Wang, B.; He, B. Predicting Herb-disease Associations Through Graph Convolutional Network. *Curr. Bioinform.* **2023**, *18*, 610–619. [CrossRef]

40. He, B.; Wang, K.; Xiang, J.; Bing, P.; Tang, M.; Tian, G.; Guo, C.; Xu, M.; Yang, J. DGHNE: Network enhancement-based method in identifying disease-causing genes through a heterogeneous biomedical network. *Brief. Bioinform*. **2022**, *23*, bbac405. [CrossRef] [PubMed]
41. Wadoux, A.M.J.-C.; Heuvelink, G.B.M.; de Bruin, S.; Brus, D.J. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecol. Model*. **2021**, *457*, 109692. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.