Supplement of Earth Syst. Sci. Data, 18, 5601–5626, 2026
https://doi.org/10.5194/essd-18-5601-2026-supplement

Supplement of

Mapping complex cropping patterns in China (2018–2021) at 10 m resolution: a data-driven framework based on multi-product integration and Google satellite embedding

Xiyu Li and Le Yu

Correspondence to: Le Yu (leyu@tsinghua.edu.cn)

Contents

S1. Data and method

S1.1 Agro-Natural Regionalization

The Agro-Natural Regionalization of China was proposed by Mr. Qiu Baojian and Mr. Huang Bingwei, which divided China into 38 agricultural natural regions based on temperature zones and moisture conditions (Figure S1, Table S1).

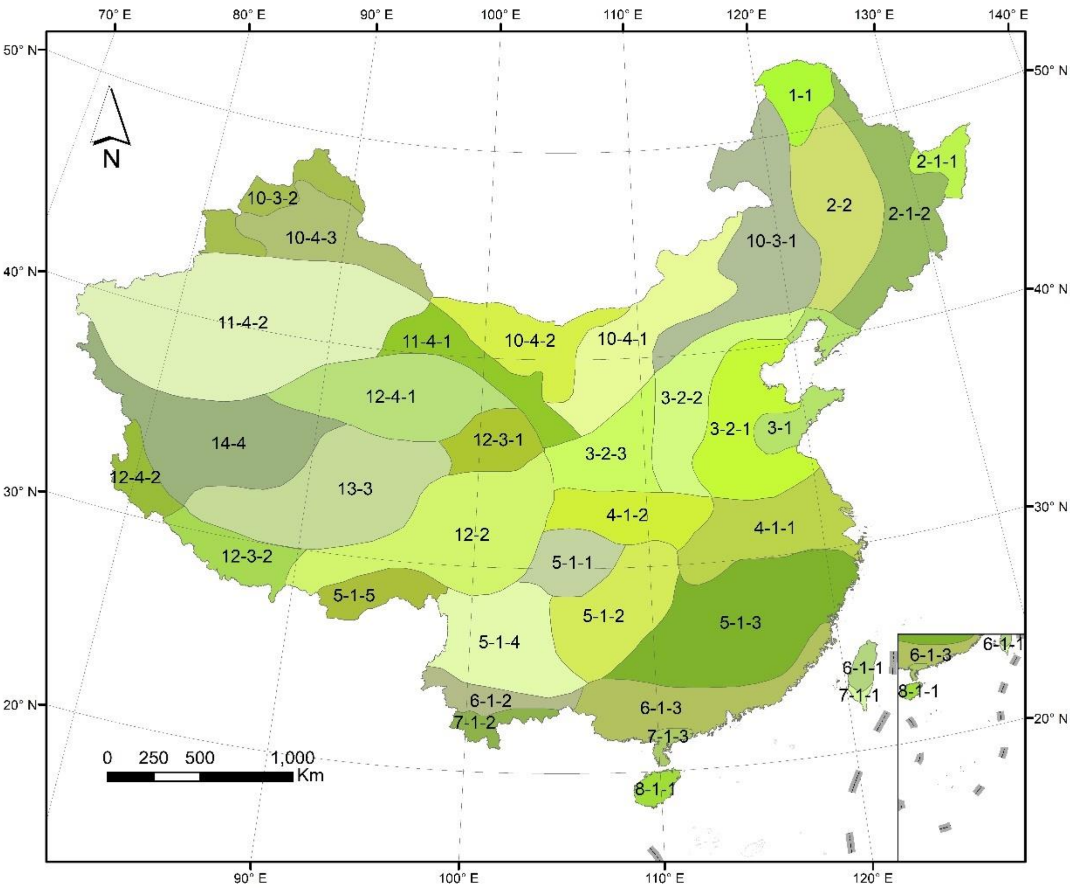

Figure S1 Agro-Natural Regionalization of China.

Based on thermal conditions, China is divided into 14 temperature zones defined by multiple indices, including accumulated temperature, mean temperature of the coldest month, and mean temperature of the warmest month (Table S2). These zones reflect the country’s thermal resources and growing season length, and they differ significantly in cropping systems and dominant agricultural products. In addition, according to moisture conditions, the whole country is further divided into four regions of different humidity regimes: humid, semi-humid, semi-arid, and arid zones (Table S3). These reflect the balance between precipitation and evapotranspiration, which directly affects vegetation patterns, land use, and agricultural practices. By overlaying the classification of temperature zones with that of moisture conditions, the framework of agricultural natural regionalization was established. This systematic division highlights how both heat (temperature) and water (moisture) jointly determine the spatial differentiation of China's agriculture, leading to distinct regional cropping systems, land-use patterns, and dominant agricultural products.

S1.2 Multi-Product coverage

We synthesize multiple publicly available crop mapping products in China (Table 1). These products include seven major crop types: maize, rice, wheat, soybean, rapeseed, sugarcane, and cotton. These products differ in spatial coverage (Table S4). Some focus on key production regions, for example, winter wheat mapping in the Huang–Huai–Hai Plain, which accounts for about 90% of China's total output, and cotton mapping in Xinjiang, which contributes more than 80% of national production. Others, such as maize and rapeseed, extend to almost all provinces nationwide. Importantly, the spatial extent of existing crop distribution products is not temporally continuous or consistent but varies across years. Some special treatments were applied to specific crops. For rice, outside the three northeastern provinces there is essentially only one dataset available, namely the double- and single-cropping rice maps developed by Yuan's research lab (Pan et al., 2021; Shen et al., 2023). In these regions, areas where two maps overlap reflect inconsistencies instead of agreement and were therefore treated separately. In addition, the soybean area in eastern Inner Mongolia, adjacent to the three northeastern provinces, was excluded from the provincial coverage since it did not represent the whole province and was treated separately as a special case.

S1.3 Consistency Analysis

In the workflow of consistency analysis, low-consistency is first identified within the same crop type based on CMCI (Crop Map Consistency Index, CMCI), reflecting inconsistencies among products of the same crop type. Subsequently, confused and unlabeled areas are determined based on the integration of multiple product types, using the relationship between OCTC (Overlapping Crop Type Counts, OCTC) and the estimated cropping intensity (e.g., OCTC > cropping intensity for confused areas, and OCTC = 0 for unlabeled areas) (Figure S2).

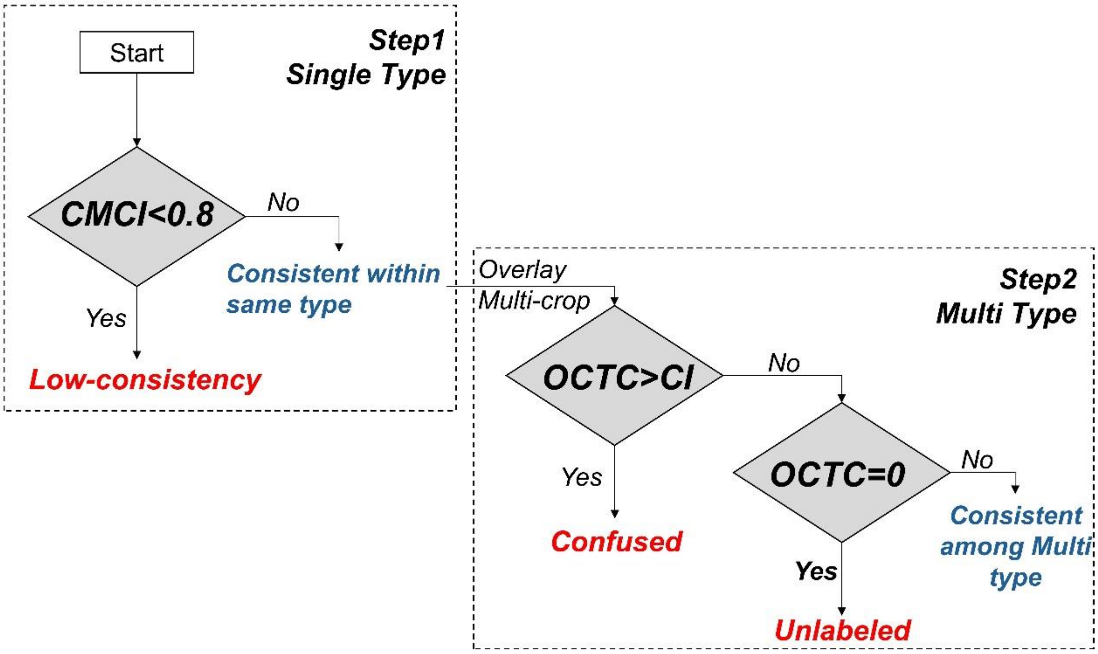

Figure S2. Workflow of Consistency Analysis

.

Table S1. Agro-Natural Regionalization of China

ID	Name	Temperature Zone	Moisture Zone
1-1	Northern Greater Khingan Range Area	Cold Temperate	Humid Zone
2-1-1	Sanjiang Plain	Middle Temperate	Humid Zone
2-1-2	Lesser Khingan–Changbai Mountain Area	Middle Temperate	Humid Zone
2-2	Songliao Plain	Middle Temperate	Semi-Humid Zone
3-1	Liaodong–Shandong Peninsula	Warm Temperate	Humid Zone
3-2-1	North China Plain	Warm Temperate	Semi-Humid Zone
3-2-2	North China Mountain Area	Warm Temperate	Semi-Humid Zone
3-2-3	Loess Plateau	Warm Temperate	Semi-Humid Zone
4-1-1	Middle–Lower Yangtze River Area	Northern Subtropical	Humid Zone
4-1-2	Upper–Middle Han River Area	Northern Subtropical	Humid Zone
5-1-1	Sichuan Basin	Central Subtropical	Humid Zone
5-1-2	Guizhou Plateau	Central Subtropical	Humid Zone
5-1-3	Jiangnan Hilly Area	Central Subtropical	Humid Zone
5-1-4	Yunnan Plateau	Central Subtropical	Humid Zone
5-1-5	Southern Slope of the Himalayas Area	Central Subtropical	Humid Zone
6-1-1	Taipei Area	Southern Subtropical	Humid Zone
6-1-2	Southern Yunnan Mountain Area	Southern Subtropical	Humid Zone
6-1-3	Fujian–Guangdong Hilly Area	Southern Subtropical	Humid Zone
7-1-1	Tainan Area	Marginal Tropical	Humid Zone
7-1-2	Valleys and Basins of Southern Yunnan	Marginal Tropical	Humid Zone
7-1-3	Qiong–Le Area (Hainan & Leizhou)	Marginal Tropical	Humid Zone
8-1-1	Southern Hainan Island Area	Mid-Tropical	Humid Zone
8-1-2	Dongsha, Zhongsha, and Xisha Islands Area	Mid-Tropical	Humid Zone

ID	Name	Temperature Zone	Moisture Zone
9-1	Nansha Islands Area	Equatorial Tropical	Humid Zone
10-3-1	Eastern Inner Mongolia Area	Arid Middle Temperate	Semi-Arid Zone
10-3-2	Northwestern Mountain Area of Northern Xinjiang	Arid Middle Temperate	Semi-Arid Zone
10-4-1	Central Inner Mongolia Area	Arid Middle Temperate	Arid Zone
10-4-2	Western Inner Mongolia Area	Arid Middle Temperate	Arid Zone
10-4-3	Northern Xinjiang Area	Arid Middle Temperate	Arid Zone
11-4-1	Hexi Corridor Area	Arid Warm Temperate	Arid Zone
11-4-2	Southern Xinjiang Area	Arid Warm Temperate	Arid Zone
12-2	Western Sichuan–Eastern Tibet Area	Plateau Temperate	Semi-Humid Zone
12-3-1	Qinghai Area	Plateau Temperate	Semi-Arid Zone
12-3-2	Southern Tibet Area	Plateau Temperate	Semi-Arid Zone
12-4-1	Qaidam Basin	Plateau Temperate	Arid Zone
12-4-2	Western Tibet Area	Plateau Temperate	Arid Zone
13-3	Southern Qiangtang Area	Plateau Sub-Frigid Zone	Semi-Arid Zone
14-4	Northern Qiangtang Area	Plateau Frigid Zone	Arid Zone

Table S2 Regional Division of Temperature Zones and Associated Cropping Systems in China

Zone Type	Zone	Accum. Temp. ≥10℃ (°C)	Coldest Month Temp. (°C)	Annual Extreme Low Mean (°C)
Monsoon	1 Cold Temperate Zone	<1700	<-30	<-45
	2 Middle Temperate Zone	1700–3500	-30 ~ -10	-45 – -25
	3 Warm Temperate Zone	3500–4500	-10 – 0	-25 – -10
	4 Northern Subtropical Zone	4500–5300	0 – 5	-10 – -5
	5 Central Subtropical Zone	5300–6500	5 – 10	-5 – 0
	6 Southern Subtropical Zone	6500–8000	10 – 15	0 – 5
	7 Marginal Tropical Zone	8000–8500	15 – 18	5 – 8
	8 Central Tropical Zone	>8500	>18	>8
	9 Equatorial Tropical Zone	>9000	>25	>20
Arid	10 Arid Middle Temperate Zone	<4000	<-10	<-20
	11 Arid Warm Temperate Zone	>4000	>-10	>-20
		Accum. Temp. ≥0℃ (°C)	Warmest Month Temp. (°C)	
High-cold	12 Plateau Frigid Zone	≤500	<6	
	13 Plateau Sub-Frigid Zone	500–1500	6 – 10	
	14 Plateau Temperate Zone	1500–3500	10 – 18	

Table S3 Regional Division of Humid and Arid Zones in China

Region	Annual Precipitation (mm)	Humidity Condition	Distribution Area	Vegetation	Land Use
Humid Zone	>800	Precipitation > Evaporation	South of the Qinling Mountains–Huaihe River line, southern Qinghai–Tibet Plateau, northeastern Inner Mongolia, eastern Northeast China	Forest	Agriculture dominated by paddy fields
Semi-Humid Zone	>400	Precipitation > Evaporation	Northeast Plain, North China Plain, most of the Loess Plateau, southeastern Qinghai–Tibet Plateau	Forest–Grassland	Agriculture dominated by dry farming
Semi-Arid Zone	<400	Precipitation < Evaporation	Inner Mongolia Plateau, part of the Loess Plateau, most of the Qinghai–Tibet Plateau	Grassland	Grassland livestock grazing, irrigated farming
Arid Zone	<200	Precipitation < Evaporation	Xinjiang, western Inner Mongolia Plateau, northwestern Qinghai–Tibet Plateau	Desert	Alpine livestock grazing, oasis irrigation farming

Table S4 Province coverage of crop type maps collected by this study

	c1	m1	m2	m3	m4	m5	m6	m7	ra1	ra2	ra3	ri1	ri2	ri3	ri4	so1	so2	so3	so4	so5	su1	su2	w1	w2	w3	w4
Anhui	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	0	1	1	1	1	1
Macau	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Beijing	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
Fujian	0	0	0	0	0	1	0	1	1	0	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0
Gansu	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
Guangdong	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0
Guangxi	0	0	0	1	0	1	1	1	1	0	1	0	0	1	1	0	0	1	0	0	1	1	0	0	0	0
Guizhou	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0
Hainan	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0
Hebei	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1
Henan	0	0	0	1	0	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	0	1	1	1	1	1
Heilongjiang	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0
Hubei	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	0	1	1	1	1	1
Hunan	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0
Jilin	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	0
Jiangsu	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	0	1	1	1	1	1
Jiangxi	0	0	0	0	0	1	0	1	1	1	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0
Liaoning	0	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	0	0	0	0	0
Inner Mongolia	0	0	0	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
Ningxia	0	0	0	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
Qinghai	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Shandong	0	0	0	1	0	1	1	1	1	0	1	0	0	0	1	0	0	1	1	1	0	1	1	1	1	1

	c1	m1	m2	m3	m4	m5	m6	m7	ra1	ra2	ra3	ri1	ri2	ri3	ri4	so1	so2	so3	so4	so5	su1	su2	w1	w2	w3	w4
Shanxi	0	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1
Shaanxi	0	0	0	1	1	1	1	1	1	0	1	0	0	0	1	0	0	1	1	0	0	1	1	1	1	1
Shanghai	0	0	0	0	0	1	0	1	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
Sichuan	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1	0	1	1	0	0	0
Taiwan	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Tianjin	0	0	0	1	0	1	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
Tibet	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Hong Kong	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Xinjiang	1	0	0	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
Yunnan	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0
Zhejiang	0	0	0	0	0	1	0	1	1	1	1	0	0	1	1	0	0	1	0	0	0	1	0	0	0	0
Chongqing	0	0	0	1	0	1	1	1	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0

c:cotton; m:maize; ra:rapeseed; ri:rice; so:soybean; su:sugarcane; w:wheat.

Products with partial rather than full provincial coverage were excluded from this table, but were given special treatment during mapping process (as described above).

S2. Results validation and comparison

S2.1 Accuracy assessment

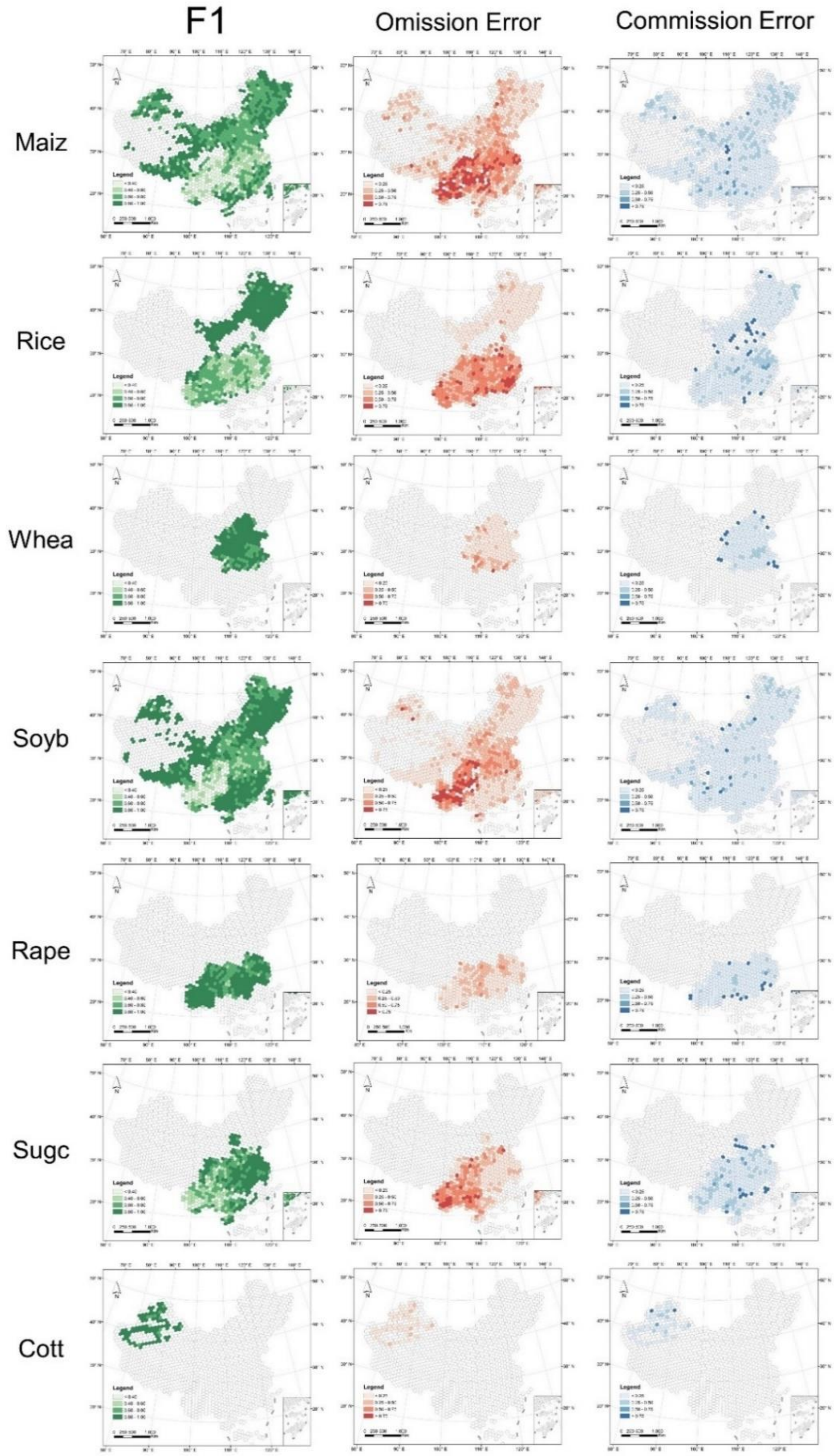

Figure S3 Spatial variation of mean accuracy of classifiers for year periods from 2018 to 2021 per crop type.

S2.2 Parameter Tuning

We conducted additional experiments to evaluate model performance under different combinations of ntree (The number of decision trees) and mtry (the number of features required for each node for splitting). To further assess the effect of parameter tuning, we computed the relative differences in OA and Kappa with respect to the baseline configuration (mtry = 8 and ntree = 500) for the year 2020.

The results show that the impact of parameter variations is generally limited. Most differences in OA are within 0.01, and differences in Kappa are similarly small, indicating that model performance is relatively insensitive to parameter tuning. For mtry, the median performance with mtry = 6 is slightly lower than the baseline, while mtry = 12 shows only marginal improvement, suggesting that the default setting (mtry = 8, the square root of the number of input features (64)) is appropriate. For ntree, performance increases initially but stabilizes around 500 trees, with differences in OA below 0.005 and in Kappa below 0.05 (Figure S4).

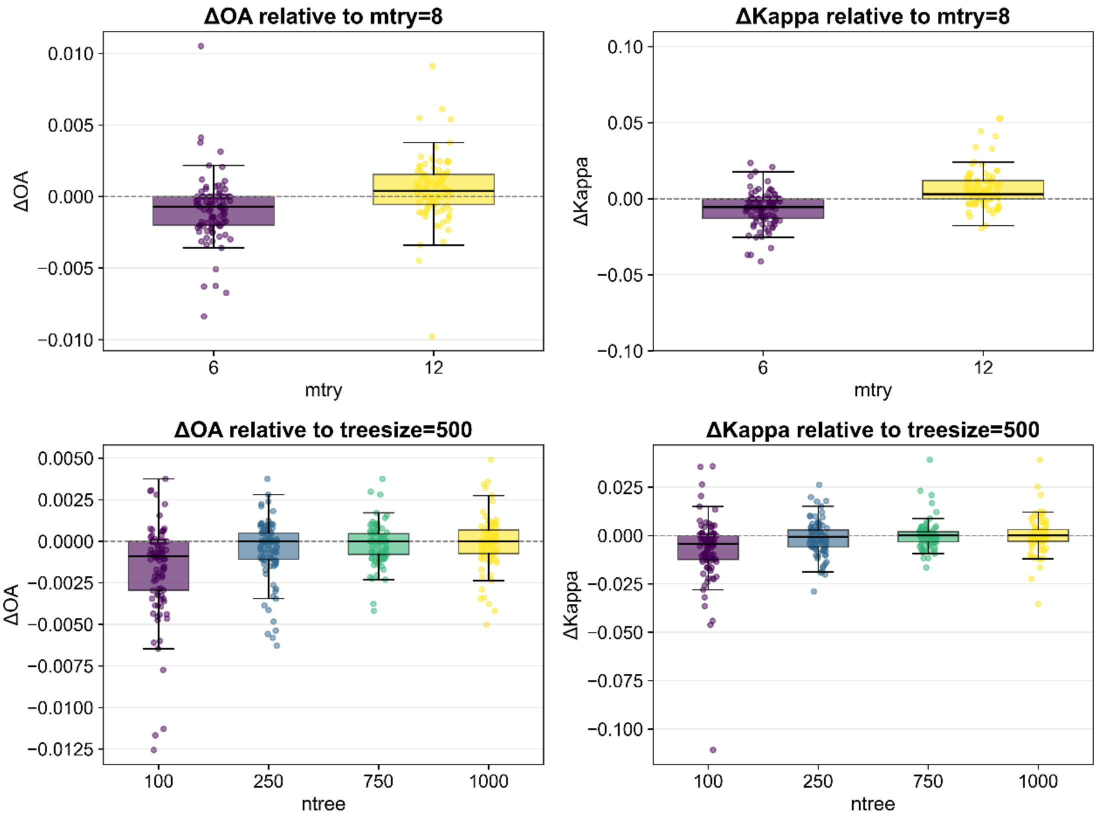

Figure S4. Relative differences in OA and Kappa compared with the baseline parameter settings.

Based on these results, we retained the baseline configuration, as it provides stable performance without unnecessary computational cost.

S2.3 Assessment of spatial autocorrelation effects on accuracy evaluation

To assess the potential influence of spatial autocorrelation on classification accuracy, we conducted an additional spatially explicit sensitivity analysis. Specifically, we assigned each sample to a hexagonal grid cell (as shown in Figure 2 in Main Text) and performed group-wise splitting at the hex level, ensuring that all samples within the same hex cell were assigned exclusively to either the training or testing set. This hex-based split reduces spatial leakage compared to random splitting.

We report the performance gap between random splitting and hex-group splitting (ΔOA and Δ kappa) as an empirical estimate of the inflation in accuracy induced by spatial autocorrelation (Figure S5). The results show that while accuracy metrics decrease under the more conservative hex-based validation, this mainly reflects the limited spatial transferability of classifiers trained on embedding features rather than deficiencies in the proposed mapping framework.

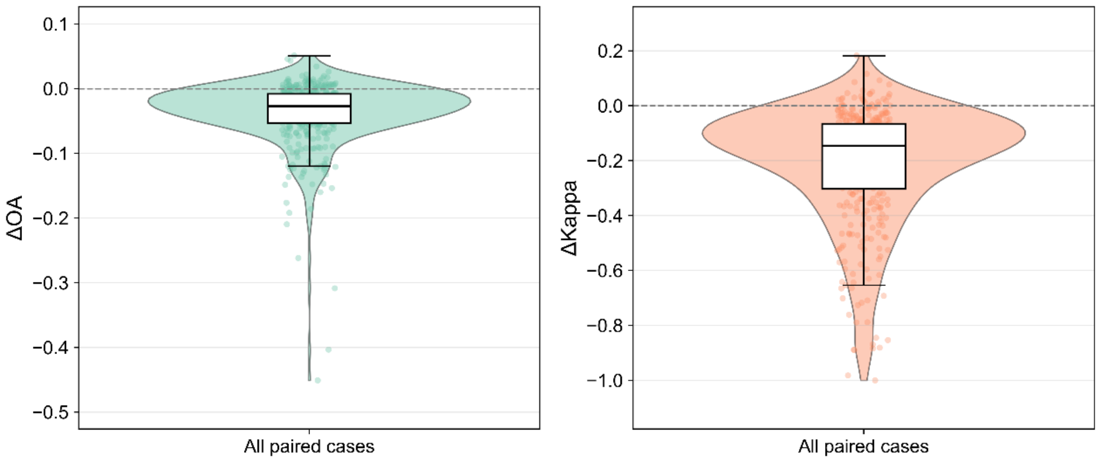

Figure S5. Relative differences in OA and Kappa under random and spatial (hex-based) splitting

It is important to note that the primary contribution of this study lies in developing a multi-product integration and reclassification framework and generating corresponding mapping products, for which the original sample-based validation strategy using random splitting provides a more representative assessment of mapping accuracy within the target regions and years. This is further supported by the fact that the training and validation samples were selected using a stratified random sampling scheme within consistent areas across multiple products, ensuring adequate spatial coverage without evident regional gaps.

S2.4 Crop-specific area comparison with existing products and Statistics

Crop-specific validation is essential for supporting practical applications, particularly for analyses targeting individual crop types. In this study, crop-specific classification accuracy, including F1-score, omission error (OE), and commission error (CE), has been systematically evaluated and is provided in Section S2 of the Supplementary Information, with corresponding discussion in Section 3.2 of the main text.

To further assess the reliability of the proposed dataset, we conducted additional crop-by-crop comparisons between our mapping results, the input crop products, and crop-specific statistical data. The results are presented in Figures S6–S12, focusing on the top five provinces for each crop in terms of planting area.

Overall, our mapping results show good consistency with both the input products and statistical data across most provinces, particularly for major staple crops such as rice, wheat, and maize. For other crops (e.g., soybean, sugarcane, and rapeseed), larger discrepancies are observed among different products and between products and statistics. In some cases, our framework helps reconcile these differences, while in others it remains more consistent with specific input products. This is because the proposed framework performs multi-product integration and reclassification at the pixel level under consistent cropland extent and cropping intensity constraints, rather than aggregating results at the provincial scale (e.g., by averaging or majority voting).

It should also be noted that this study includes only seven single-cropping types and five major crop sequences, while all other cropping patterns are grouped into an “other multi-cropping” category. As a result, some cropping systems involving these crops are partially absorbed into this aggregated class, which may lead to systematic underestimation of crop-specific areas. This effect is particularly evident for rice, where both our mapping results and similar products tend to be lower than statistical data, and is also observed for other crops such as rapeseed.

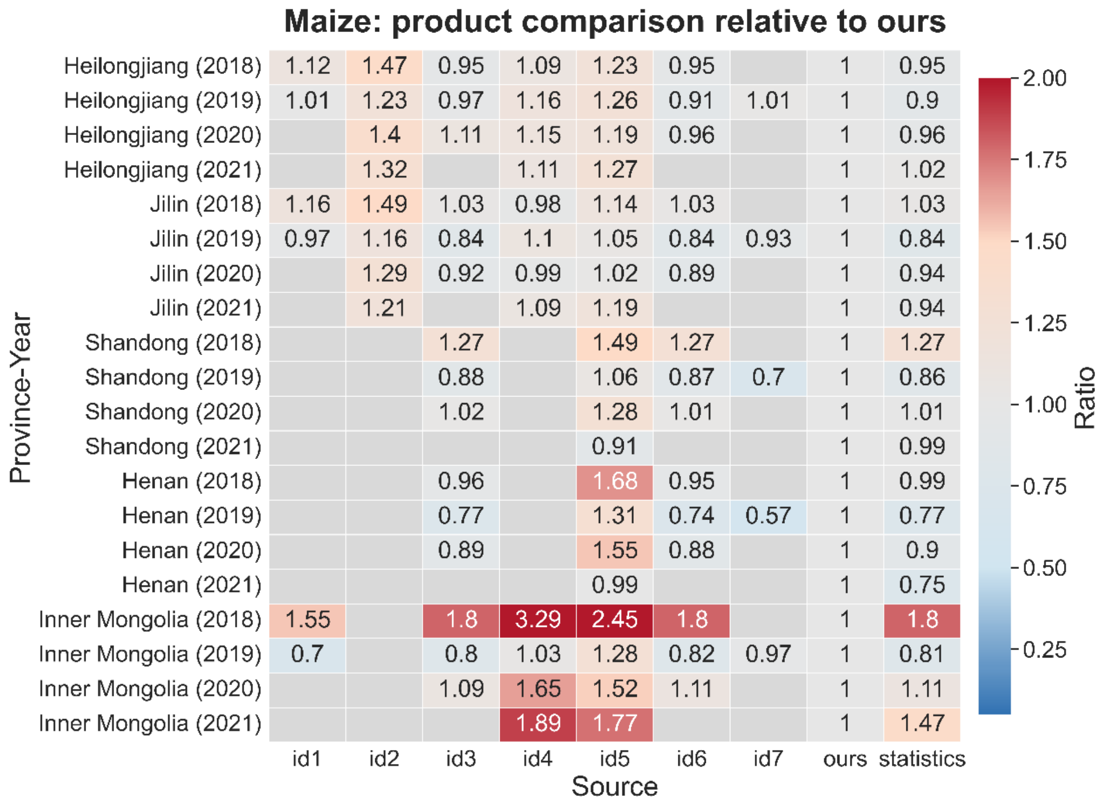

Figure S6. Comparison of maize area among our mapping results, input products, and statistics.

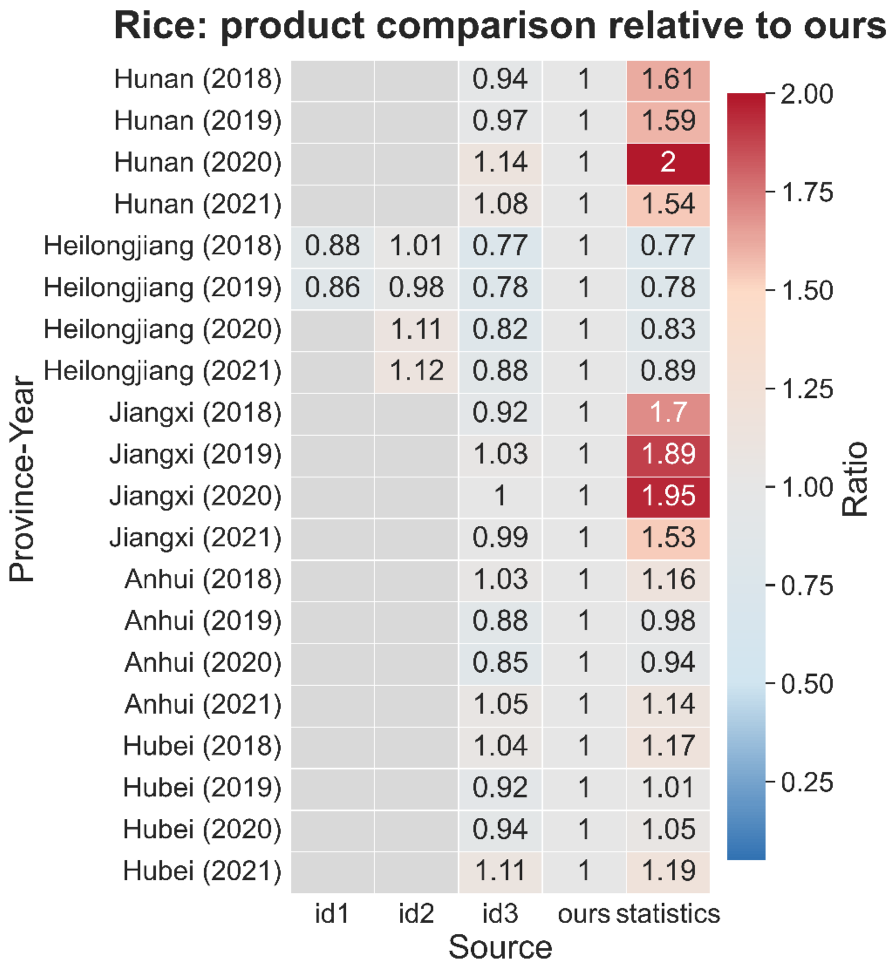

Figure S7. Comparison of rice area among our mapping results, input products, and statistics. The reported area of product id3 corresponds to the combined area of single- and double-season rice (i.e., id3 + id4).

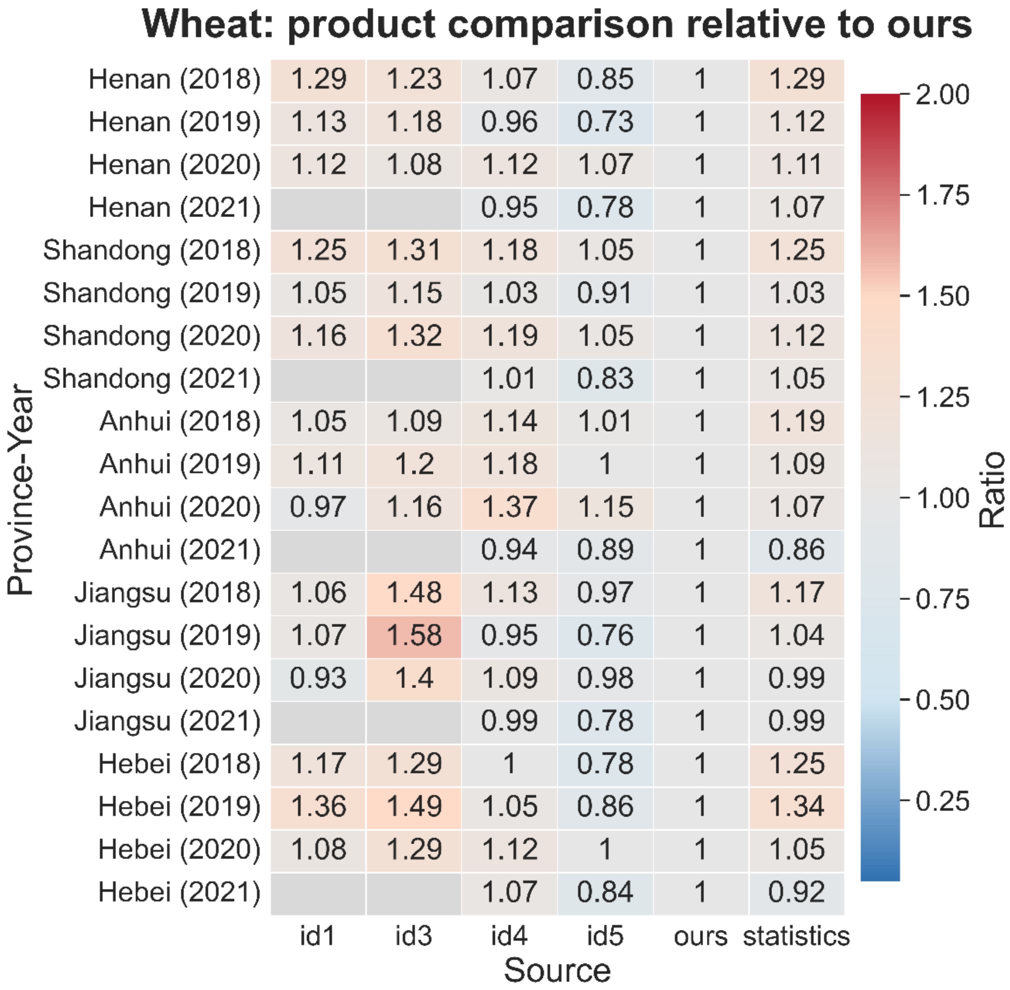

Figure S8. Comparison of wheat area among our mapping results, input products, and statistics.

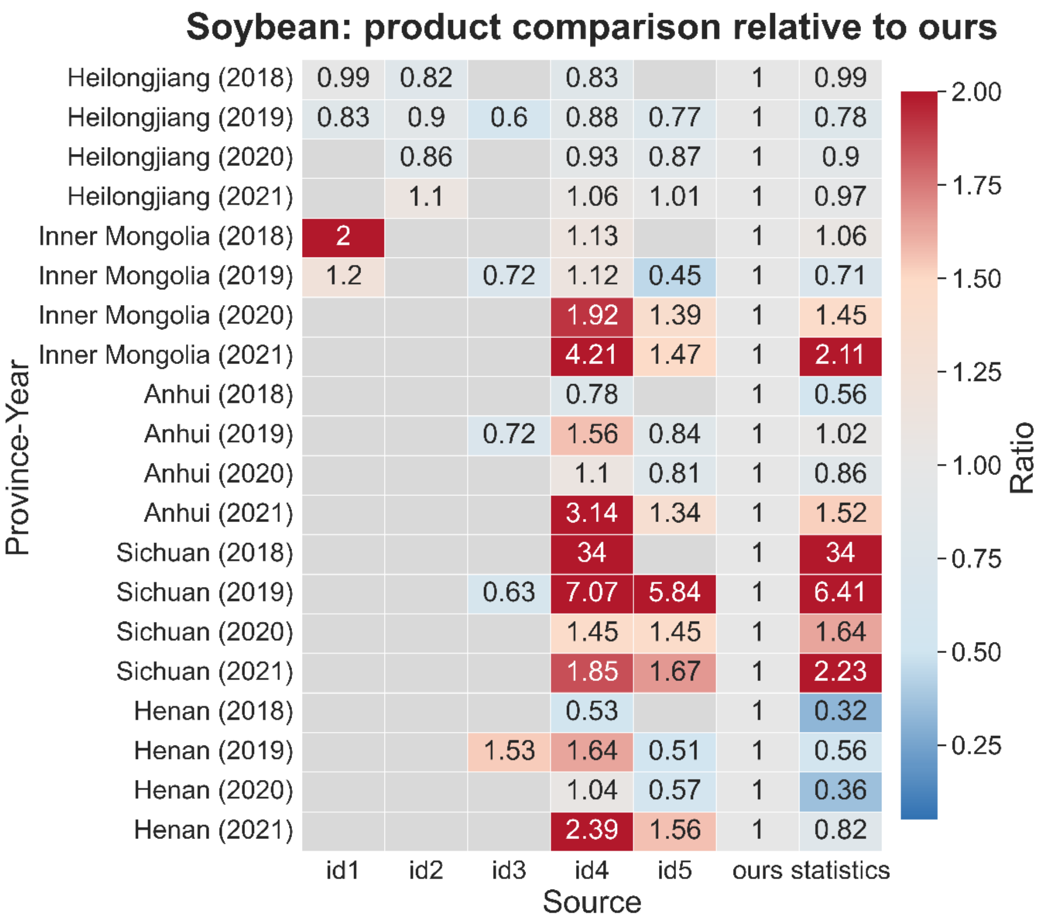

Figure S9. Comparison of soybean area among our mapping results, input products, and statistics.

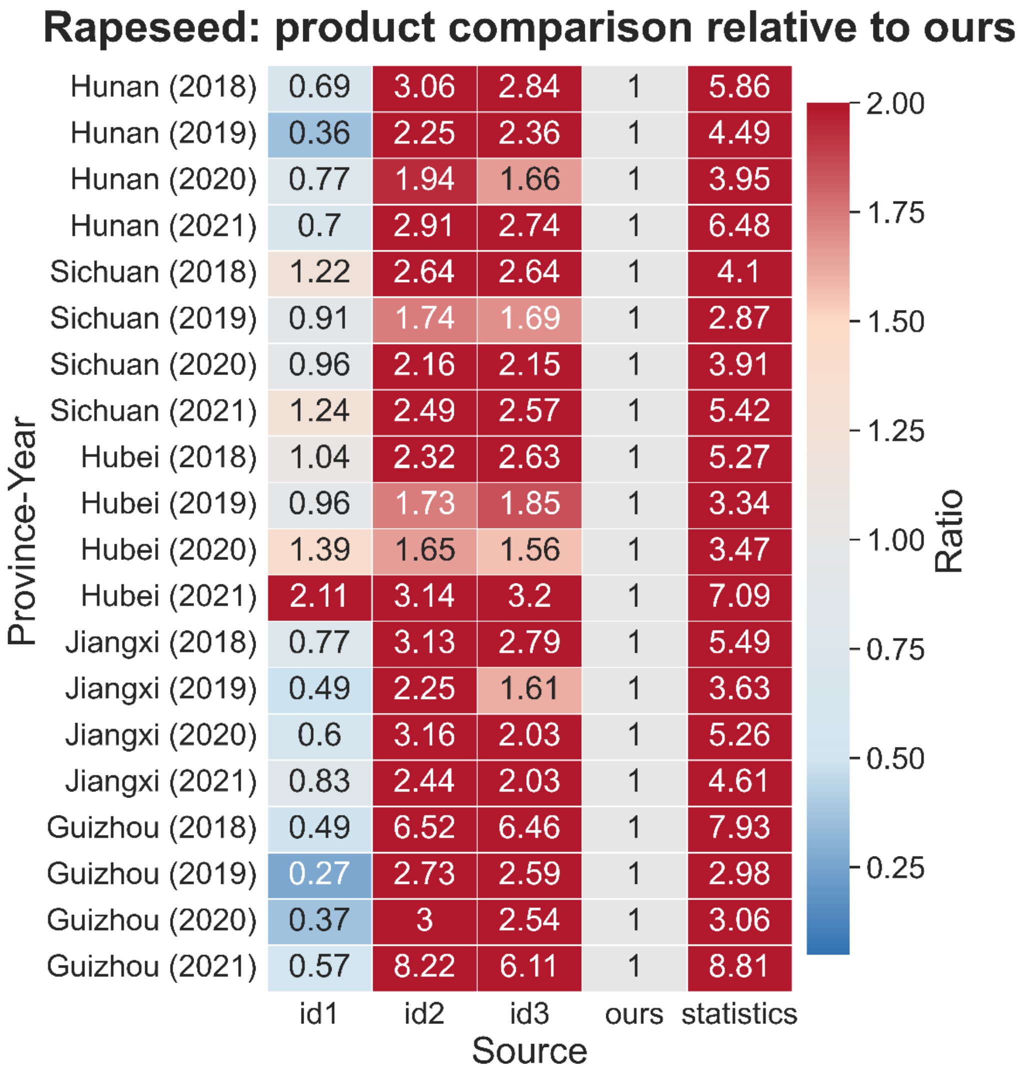

Figure S10. Comparison of rapeseed area among our mapping results, input products, and statistics.

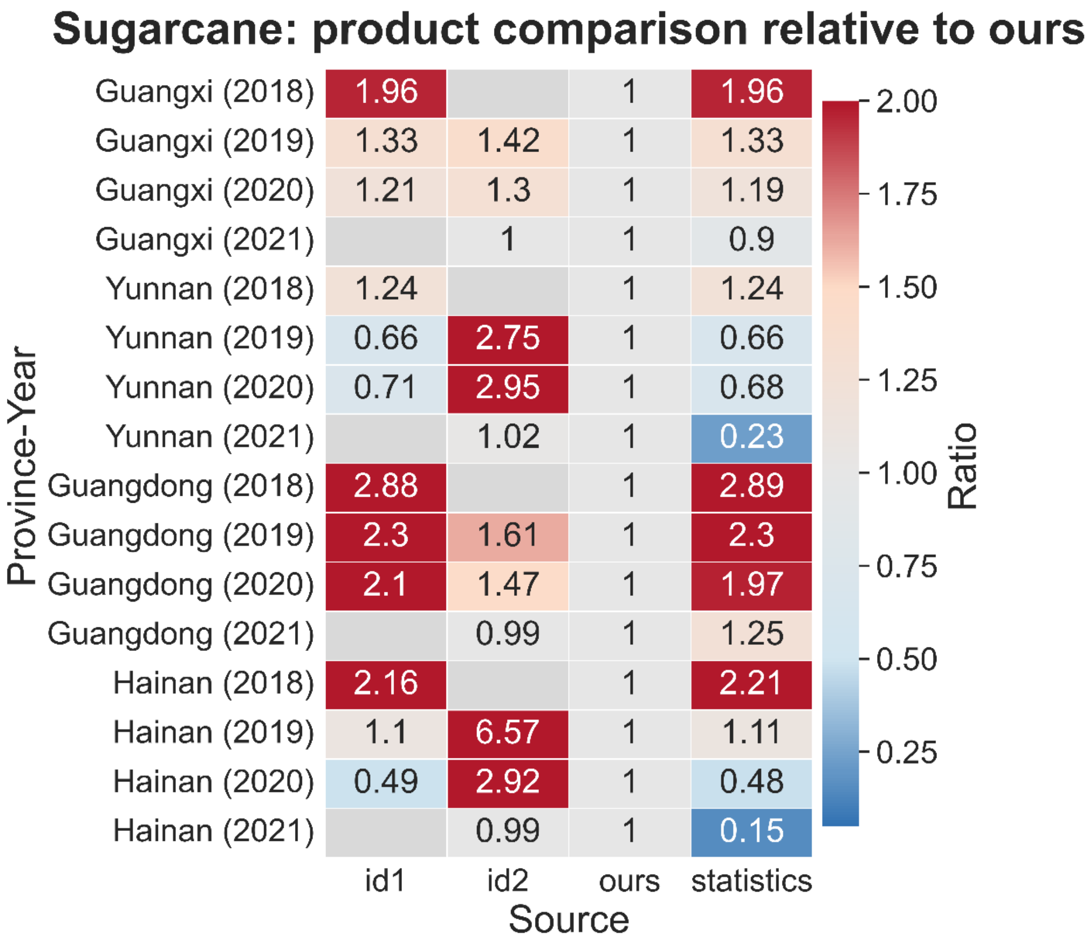

Figure S11. Comparison of sugarcane area among our mapping results, input products, and statistics.

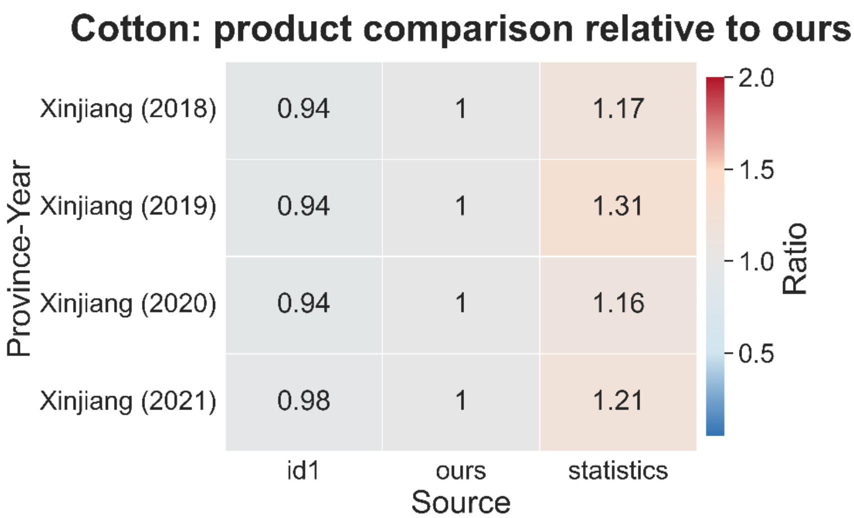

Figure S12. Comparison of cotton area among our mapping results, input products, and statistics.

S2.5 Interpretability analysis of Google Satellite Embedding

Although individual embedding dimensions do not have explicit physical meanings, recent studies suggest that they encode complex combinations of spectral, structural, and phenological information derived from multi-sensor observations (Alam and Simic Milas, 2025).

To further enhance interpretability, we conducted an additional analysis linking embedding features to raw remote sensing signals. Specifically, instead of using a large set of derived indices, we retained only the original optical (Sentinel-2 bands B2–B12) and SAR features (Sentinel-1 VV and VH) to reduce collinearity and improve physical interpretability. Cloud-contaminated pixels in Sentinel-2 imagery were masked using scene classification and cloud probability thresholds (60%). Both optical and SAR observations were aggregated into 10-day median composites, and missing values were filled using linear interpolation to ensure consistent and continuous time series.

For each crop type, we extracted embedding features and corresponding multi-temporal raw-band observations, and computed correlations between embedding features and raw bands across time, resulting in a time × band correlation matrix (Figure S13-15). Signed correlation heatmaps were generated to characterize the direction and strength of these relationships.

To facilitate interpretation, we selected three representative crops in the Northeast China Plain (maize, rice, and soybean) and visualized the most important embedding dimensions identified in their classification models. The temporal axis represents10-day composited time steps throughout the growing season (DOY 90-300).

The results reveal crop-specific associations between embedding features and raw spectral signals. For maize, the most important embedding feature (A08) shows strong negative correlations with shortwave infrared bands (B11–B12) during DOY ~200–270, indicating sensitivity to canopy moisture dynamics during the mid-to-late growing season. For rice, A40 exhibits strong positive correlations with red-edge and SWIR bands (B6–B12) during DOY ~110–130, corresponding to the transplanting and early growth stages, where distinct water–soil conditions enhance separability from upland crops. For soybean, A33 shows strong positive correlations with red-edge bands (B6–B8A) during DOY ~220–270 and with SWIR bands (B11–B12) during DOY ~280–300, reflecting canopy development and subsequent moisture decline during senescence.

These patterns are consistent with known crop phenology and spectral separability in the region, where SWIR and red-edge bands play key roles in distinguishing crop types at critical growth stages. Compared with interpretations based on high-dimensional derived features, this analysis provides a more direct and physically interpretable understanding of embedding representations.

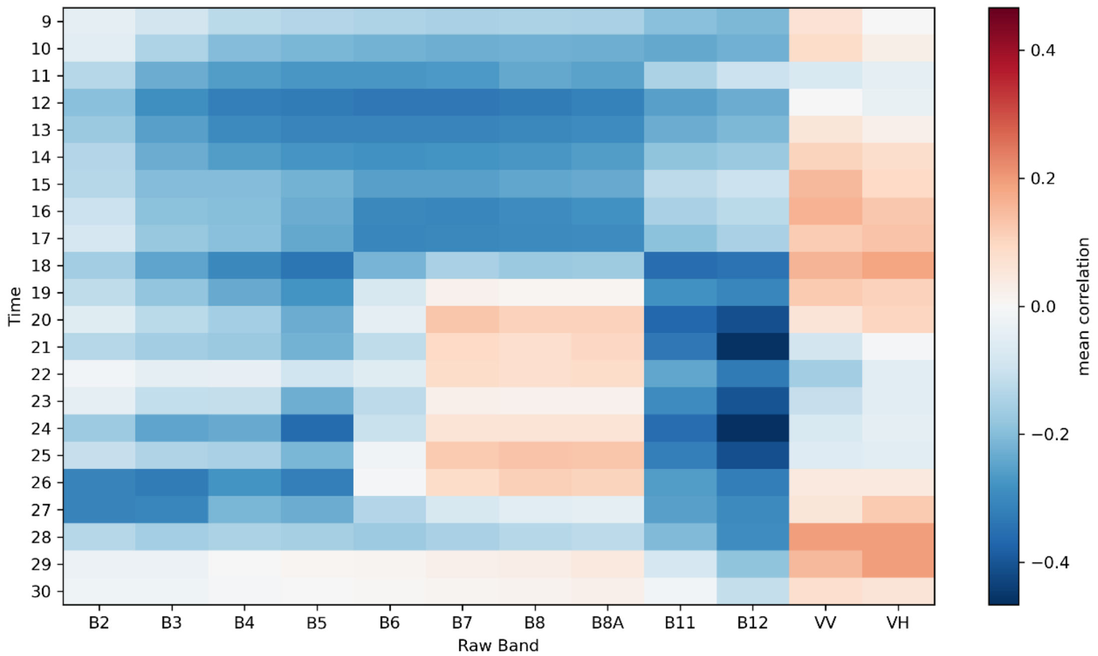

Figure S13. Correlation heatmap between embedding feature A08 and multi-temporal bands (Maize in Northeast China)

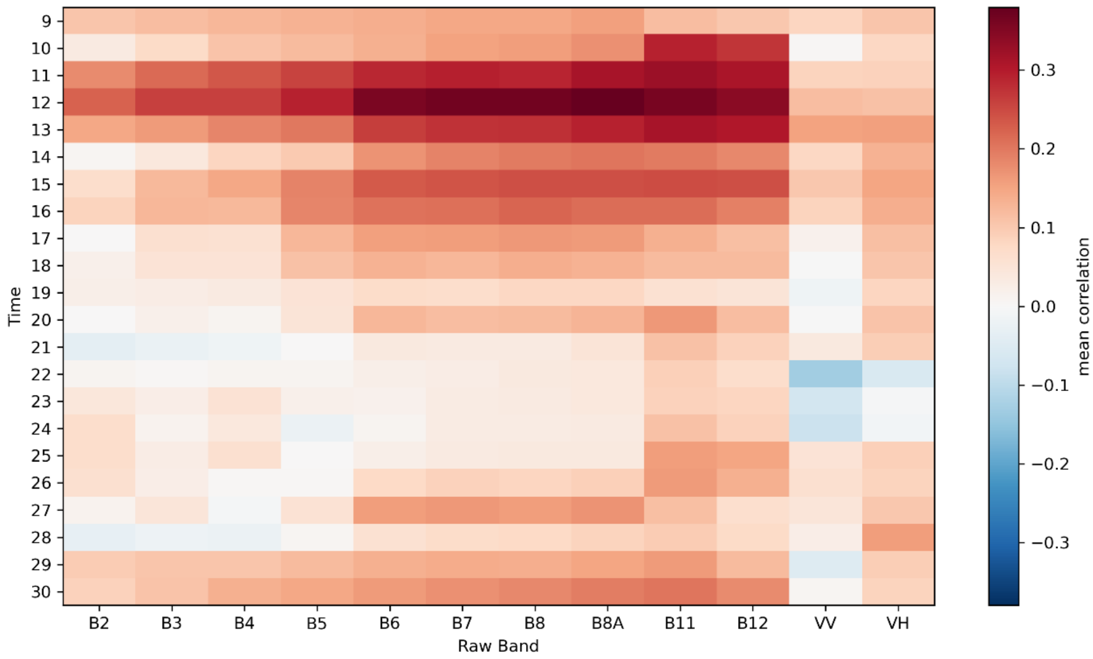

Figure S14. Correlation heatmap between embedding feature A40 and multi-temporal bands (Rice in Northeast China)

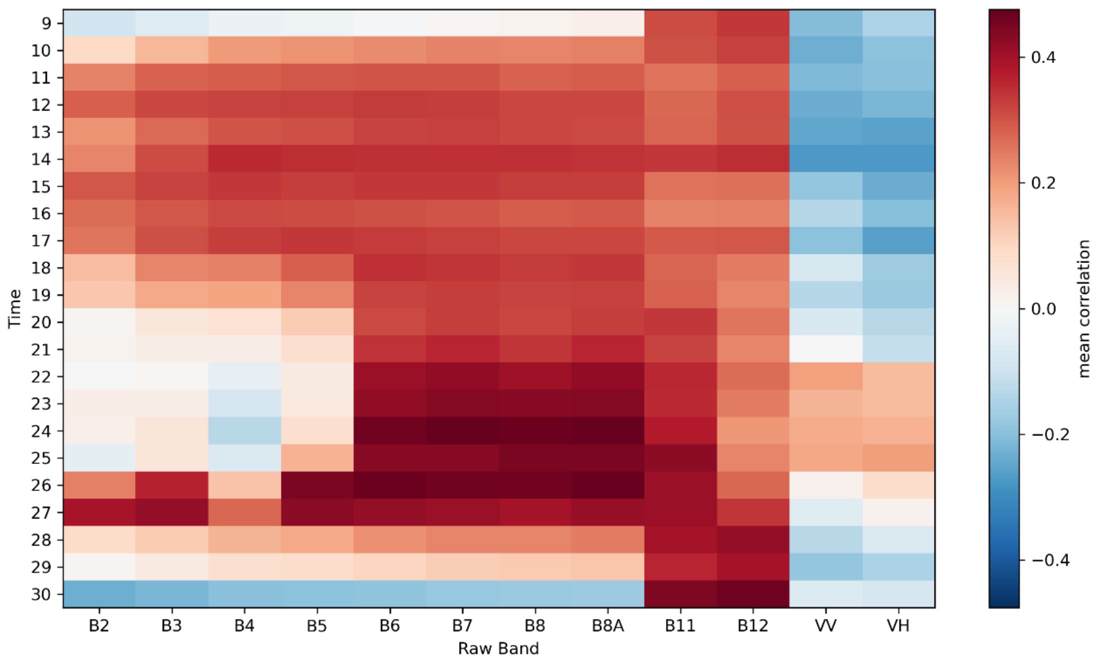

Figure S15. Correlation heatmap between embedding feature A33 and multi-temporal bands (Soybean in Northeast China)

Reference:

Huang, B.W. (1959). Draft of China's comprehensive physical regionalization. Science Bulletin, 18, 594–602. [in Chinese]

Qiu, B.J. (1989). Agricultural natural regionalization. In National Agricultural Atlas (pp. 20–21). China Cartographic Publishing House. [in Chinese]

Pan, B., Zheng, Y., Shen, R., Ye, T., Zhao, W., Dong, J., . . . Yuan, W. (2021). High Resolution Distribution Dataset of Double-Season Paddy Rice in China. Remote Sensing, 13(22). doi:10.3390/rs13224609

Shen, R., Pan, B., Peng, Q., Dong, J., Chen, X., Zhang, X., . . . Yuan, W. (2023). High-resolution distribution maps of single-season rice in China from 2017 to 2022. Earth Syst. Sci. Data, 15(7), 3203-3222. doi:10.5194/essd-15-3203-2023

Alam, M. M. T. and Simic Milas, A.: Dimensionality optimized machine learning retrieval of canopy chlorophyll, nitrogen, and phosphorus from google satellite embeddings, Smart Agricultural Technology, 12, 101601, https://doi.org/10.1016/j.atech.2025.101601, 2025.