Responses to editor and reviewers

We sincerely appreciate the constructive comments which led us to improve the quality of this manuscript (ESSD-2025-326: CN_Wheat10: A 10 m resolution dataset of spring and winter wheat distribution in China (2018–2024) derived from time-series remote sensing). The responses to suggestions and comments are shown in **blue** text. We have highlighted the revised sections and corresponding references in **red** text. The page (P) and line (L) numbers indicated in the response refer to the revised manuscript. The item-by-item responses to all comments are listed below.

Reviewer #2

This paper has been improved based on the previous comments. However, several issues still need to be addressed before publication.

Response: We sincerely thank the reviewer for positive assessment that our manuscript has been improved and for continued engagement in providing further constructive feedback. We have carefully studied all points and have made revisions to the manuscript to address them. Our point-by-point responses to each specific comment are detailed below.

Specific comments:

Point 1. A more explicit workflow of the classification procedure is needed. The existing Figure 2 focuses primarily on feature selection and fails to convey the complete process. The overall workflow is difficult to follow, particularly given the updated classification steps involving the predefined spring- and winter-wheat-dominated provinces and the temporal windows applied to different cases.

Response: Thank you for this valuable comment. We agree that the previous workflow diagram did not sufficiently illustrate the complete classification process. In the revised manuscript, we have updated Fig.R1 to present a clearer and more comprehensive workflow. Specifically, we have incorporated the predefined spring- and winter-wheat-dominated provinces into the sample generation stage, ensuring that the classification logic related to different wheat-growing regions is explicitly shown. In addition, we have added schematic illustrations within the feature-selection section to clarify the distinct feature-selection strategies for spring and winter wheat. These revisions make the entire workflow easier to follow and more accurately reflect the updated classification steps described in the manuscript.

The corresponding content has been made on Line 205 in Page 8 of the revised manuscript.

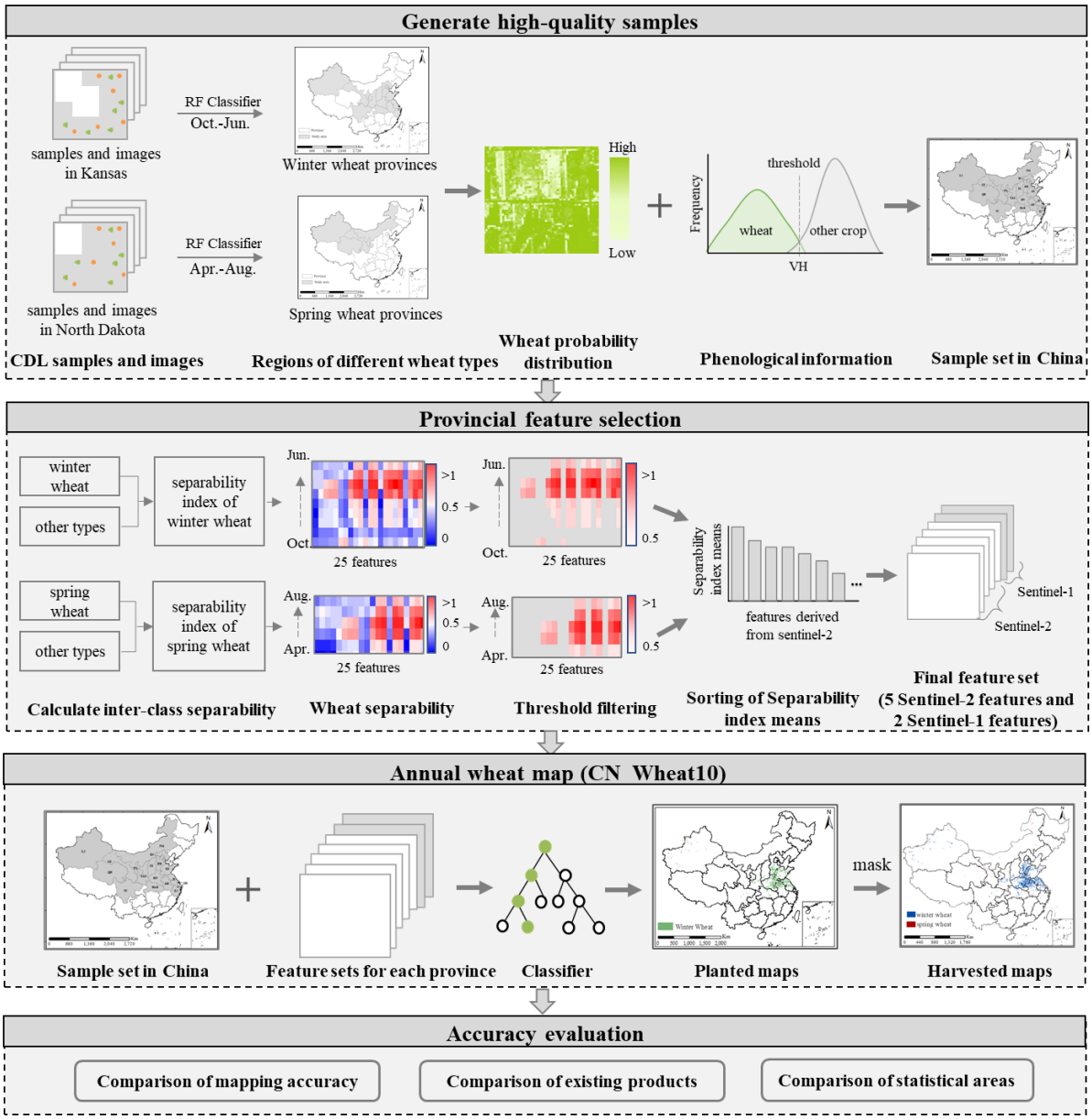

Figure R1. Flowchart for mapping annual wheat distribution.

Point 2. The validation section (Figures 8, 9, and 10) lacks consistency. The comparison with WorldCereal is presented only visually, without any quantitative accuracy assessment. The same issue applies to the planted and harvested maps from ChinaWheatMap10, which were compared visually but not evaluated for accuracy in Figure 10. It should also be noted that WorldCereal includes both spring and winter cereals, whereas Table 1 and the validation focus solely on spring cereals.

Response: We thank the reviewer for the thoughtful and constructive comments. In the revised manuscript, we have improved the consistency of the validation section (Fig. R2-R4). Specifically, we have updated all three figures so that they now include every product used for comparison. We have also added quantitative accuracy assessments for both WorldCereal and the planted/harvested maps from ChinaWheatMap10, addressing the reviewer's concern about the previous reliance on visual comparison only. In addition, we

have incorporated winter cereals from WorldCereal into Table R1 and the validation analysis, ensuring that the products listed in Table R1 are fully aligned with those included in Fig.R2-R4. These revisions make the validation more consistent, comprehensive, and comparable across all datasets.

The corresponding content has been made in Page 16 of the revised manuscript.

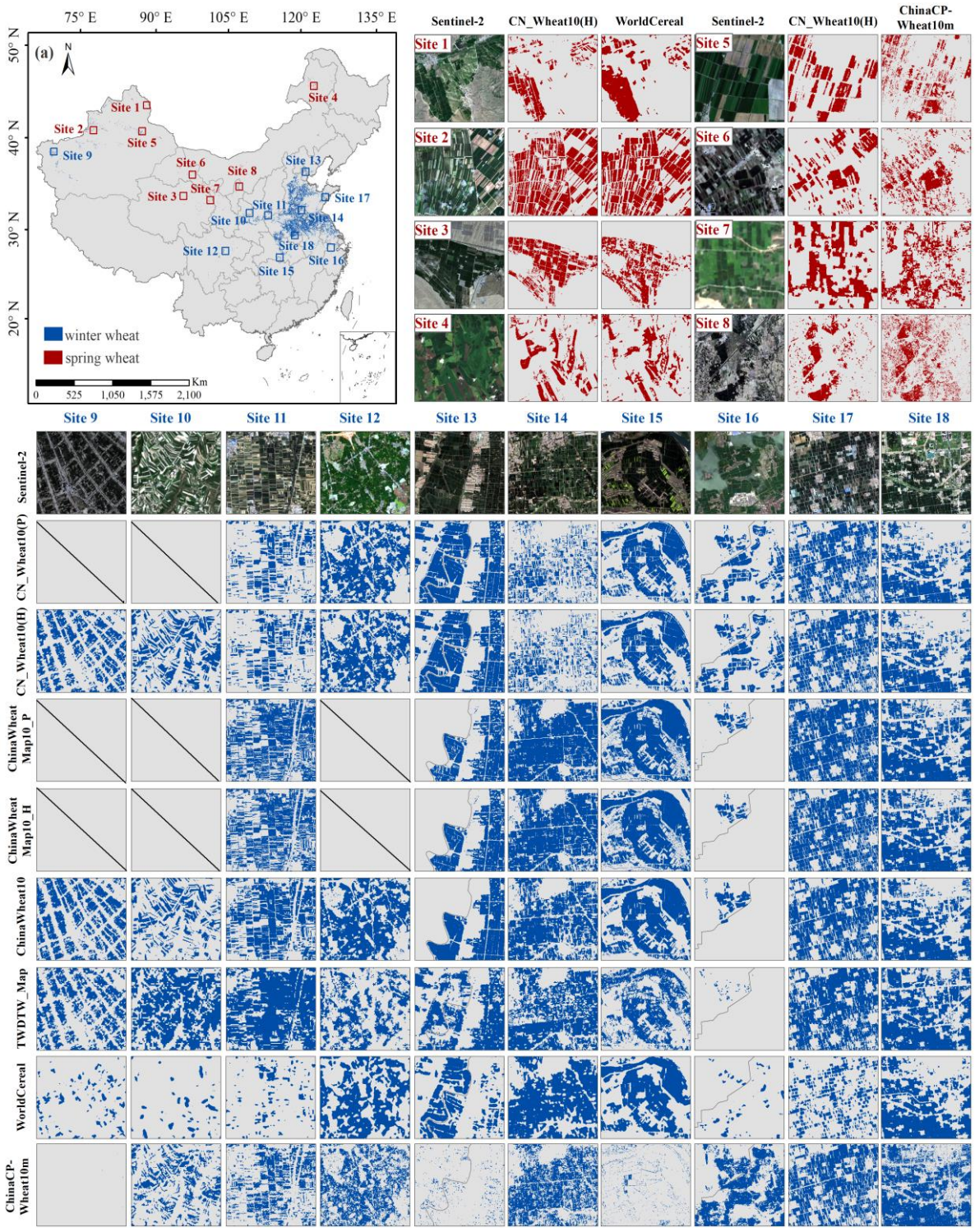

Figure R2: Comparison of wheat details between CN_Wheat10 products and existing published products.

The corresponding content has been made in Page 17 of the revised manuscript.

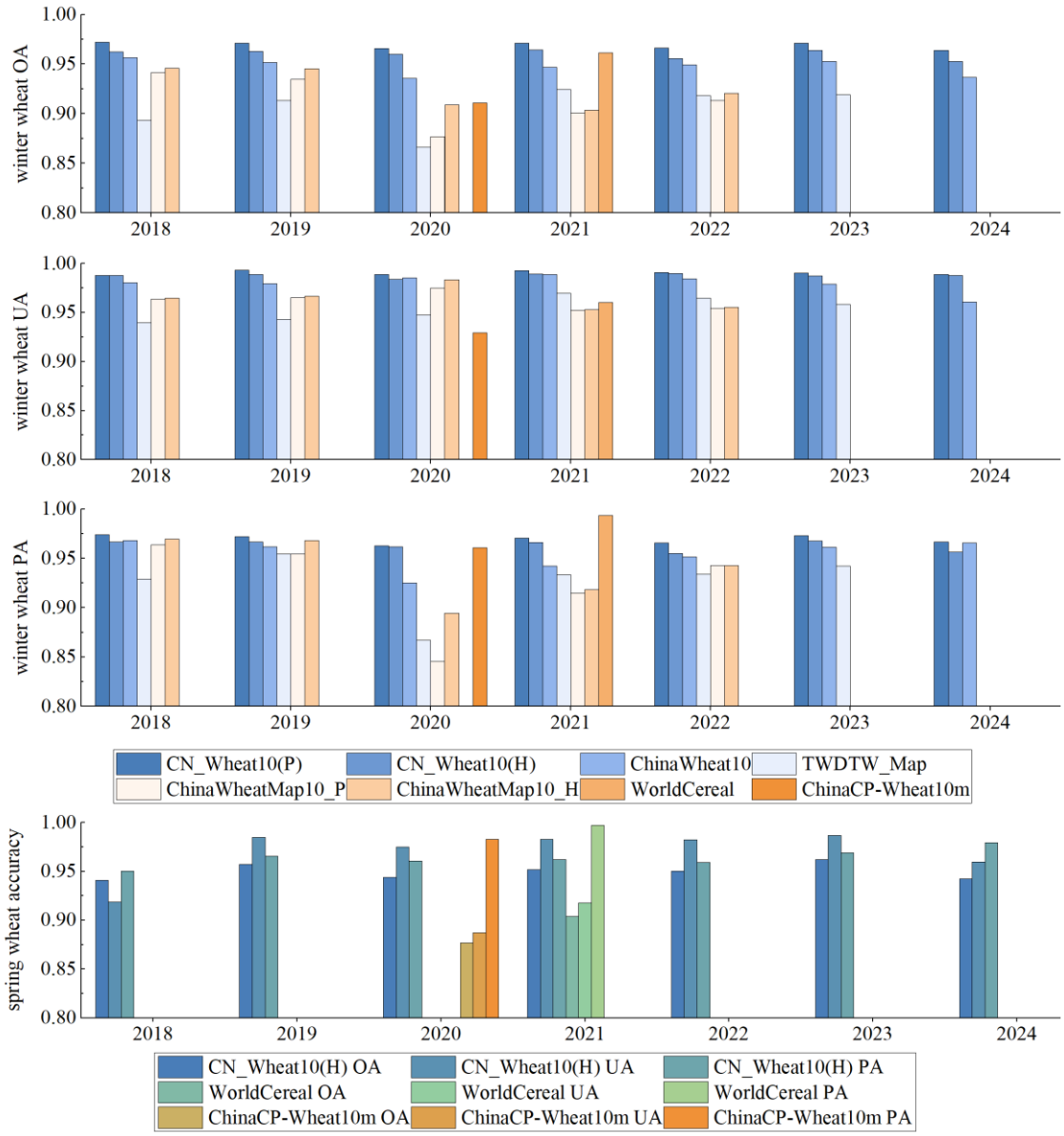

Figure R3: The mapping accuracy for spring and winter wheat from 2018 to 2024.

The corresponding content has been made in Page 18-19 of the revised manuscript.

The Fig.R4 shows that ChinaCP-Wheat10m and WorldCereal achieve relatively high accuracies in certain provinces. This is largely because both products provide wheat distribution maps for a single year, reflecting the accuracy for that specific year only, whereas the other products report multi-year average accuracies.

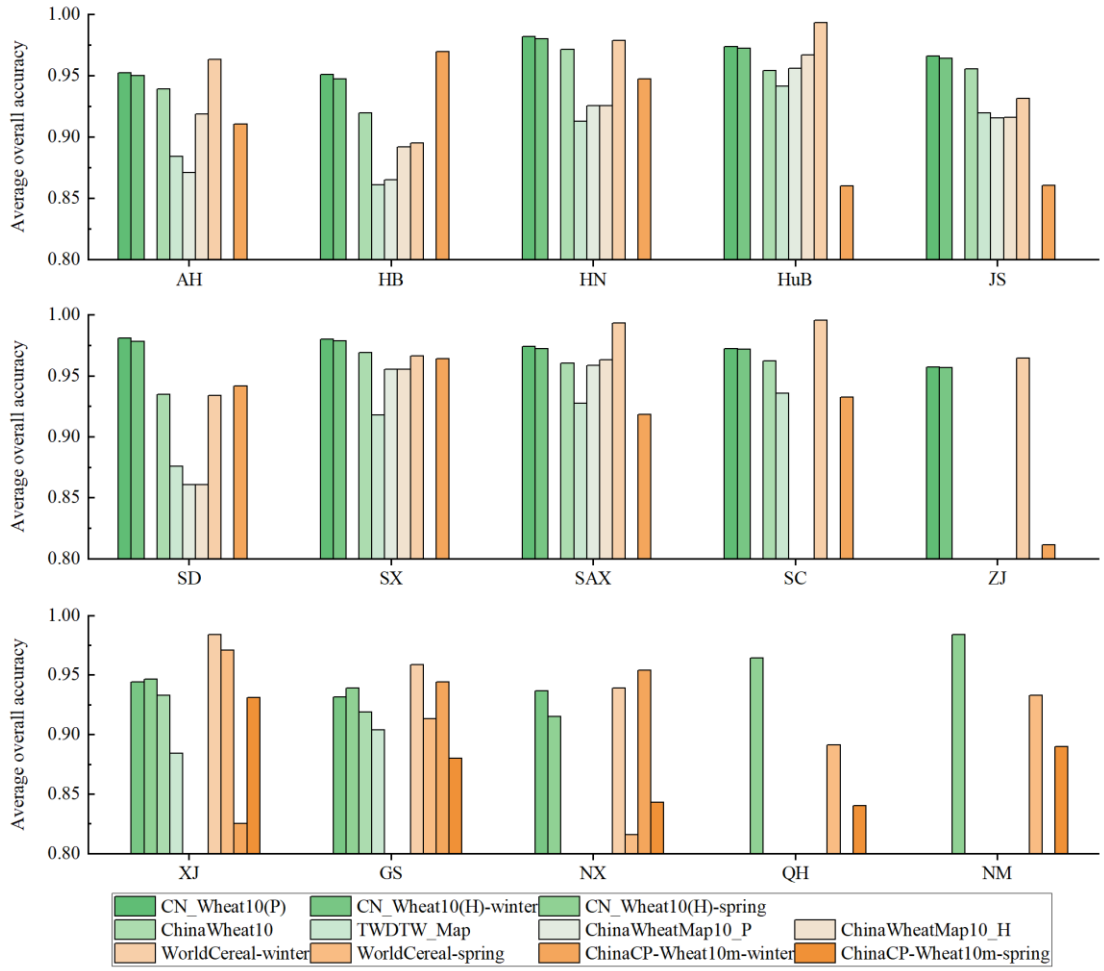

Figure R4: The average overall accuracy of wheat at the provincial level from 2018 to 2024.

The corresponding content has been made in Page 7 of the revised manuscript.

Table R1: Information on the reference wheat mapping products used for comparison.

Wheat maps	Wheat types	Study area	Resolution	Time range	Reference
ChinaWheat10	winter wheat	11 provinces	10 m	2018-2024	(Yang et al., 2023)
ChinaWheatMap10	winter wheat	8 provinces	10 m	2018-2022	(Hu et al., 2024)
ChinaCP-Wheat10m	spring and winter wheat	China	10 m	2020	(Qiu et al., 2025)
WorldCereal	spring and winter cereals	Global	10 m	2021	(Van Tricht et al., 2023)
TWDTW_Map	winter wheat	11 provinces	30 m	2001-2023	(Dong et al., 2020)

Note: ChinaWheatMap10 includes planted area maps (ChinaWheatMap10_P) and harvested area maps (Chinawheatmap10_H). The last product was generated by TWDTW algorithm, we call this product TWDTW_Map for short.

Point 3. Area estimation should be based on statistical approaches rather than simple pixel counting, and the associated uncertainty needs to be discussed in Section 4.2 & 4.3.

(See CEOS WGCV LPV Land Cover Validation Protocol, 2025. Available at: https://lpvs.gsfc.nasa.gov/PDF/CEOS_WGCV_LPV_Land_Cover_protocol_Sept2025_V1.pdf)

Response: We thank the reviewer for this constructive comment. In the revised manuscript, we performed calculations using a statistically based area estimation method following good practice recommended by Olofsson et al.(2014). Specifically, we now estimate class areas using sample-based estimators derived from the confusion matrix, and report the associated standard errors and 95 percent confidence intervals. We have also added a detailed discussion of area estimation uncertainty in Sections 4.2 and 4.3. These revisions strengthen the rigor of the area estimation procedure and address the reviewer's concern.

P19-20, L383-395: The annual wheat areas and their associated uncertainties for the period 2018–2024 were estimated using the method recommended by existing studies (Olofsson et al., 2014), with key results summarized in Table 2. For the ten major wheat provinces, the estimated planted area ranged from 16,680.34 ± 431.36 thousand hectares to 19,702.84 ± 277.17 thousand hectares, accounting for an average of approximately 10.6% of the total area. The estimated harvested area in these provinces ranged from 16,121.57 ± 561.61 thousand hectares to 19,130.85 ± 407.08 thousand hectares, averaging about 10.14% of the total. For the fifteen wheat provinces, the estimated harvested area was between 20,362.92 ± 1,513.43 and 24,684.57 ± 2,049.75 thousand hectares, with a mean proportion of approximately 3.91%. Overall, the annual wheat maps demonstrate good performance, as reflected in the 95% confidence intervals of the area estimates. It is important to note that the relatively higher uncertainty in the harvested area estimates is primarily attributed to the class imbalance in the stratified validation samples. This does not necessarily indicate poor map accuracy but reflects a statistical limitation of the sampling design, a phenomenon documented in previous studies (Liu and Zhang, 2023; Yadav and Congalton, 2017).

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sens. Environ., 148, 42-57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

Liu, W. and Zhang, H.: Mapping annual 10 m rapeseed extent using multisource data in the Yangtze River Economic Belt of China (2017–2021) on Google Earth Engine, Int. J. Appl. Earth Obs. Geoinf., 117, 103198, https://doi.org/10.1016/j.jag.2023.103198, 2023.

Yadav, K. and Congalton, R. G.: Issues with large area thematic accuracy assessment for mapping cropland extent: a tale of three continents, Remote Sens., 10, 53, https://doi.org/10.3390/rs10010053, 2017.

Table R2: Error-adjusted area estimates for annual CN_Wheat10 from 2018 to 2024.

Year	Attribute/Strata	planted area (10 provinces)	harvested area (10 provinces)	harvested area (15 provinces)
2018	Area proportion (%)	10.18	9.58	3.86
	Standard error ($\times 10^3$ ha)	244.24	270.53	752.89
	Estimated area ($\times 10^3$ ha)	17607.72	16570.03	21980.48
	95% CI in ± ($\times 10^3$ ha)	478.71	530.25	1475.67
2019	Area proportion (%)	9.64	9.32	3.57
	Standard error ($\times 10^3$ ha)	220.08	286.54	772.16
	Estimated area ($\times 10^3$ ha)	16680.34	16121.57	20362.92
	95% CI in ± ($\times 10^3$ ha)	431.36	561.61	1513.43
2020	Area proportion (%)	10.16	9.81	3.96
	Standard error ($\times 10^3$ ha)	155.55	223.10	715.43
	Estimated area ($\times 10^3$ ha)	17583.70	16966.48	22552.06
	95% CI in ± ($\times 10^3$ ha)	304.88	437.27	1402.24
2021	Area proportion (%)	11.39	11.06	4.15
	Standard error ($\times 10^3$ ha)	141.41	207.69	700.33
	Estimated area ($\times 10^3$ ha)	19702.84	19130.85	23654.27
	95% CI in ± ($\times 10^3$ ha)	277.17	407.08	1372.66
2022	Area proportion (%)	10.73	10.26	3.84
	Standard error ($\times 10^3$ ha)	162.02	227.36	745.15
	Estimated area ($\times 10^3$ ha)	18560.11	17752.06	21863.52
	95% CI in ± ($\times 10^3$ ha)	317.56	445.63	1460.50
2023	Area proportion (%)	11.09	10.46	4.33
	Standard error ($\times 10^3$ ha)	317.15	369.83	1045.79
	Estimated area ($\times 10^3$ ha)	19181.10	18096.78	24684.57
	95% CI in ± ($\times 10^3$ ha)	621.62	724.86	2049.75
2024	Area proportion (%)	11.01	10.53	3.69
	Standard error ($\times 10^3$ ha)	269.37	289.78	650.31
	Estimated area ($\times 10^3$ ha)	19042.65	18219.26	21003.42
	95% CI in ± ($\times 10^3$ ha)	527.97	567.97	1274.61

P23-24, L427-433: Since both the planted and harvested area estimates are derived from statistically based area-estimation methods recommended in previous studies, each estimate is associated with its own standard error. Accordingly, the error of the area difference should be regarded as a combination of the uncertainties from the two area estimates, and its magnitude can be quantified using standard error-propagation formulas. Based on the calculations using Table R2, the annual uncertainty range of the area differences across the ten provinces is approximately 251.26 to 487.19 thousand hectares. In addition, the imbalance in class proportions among the validation samples may further increase the uncertainty of area estimation, thereby affecting the stability of the area differences.

P25, L436-L439:

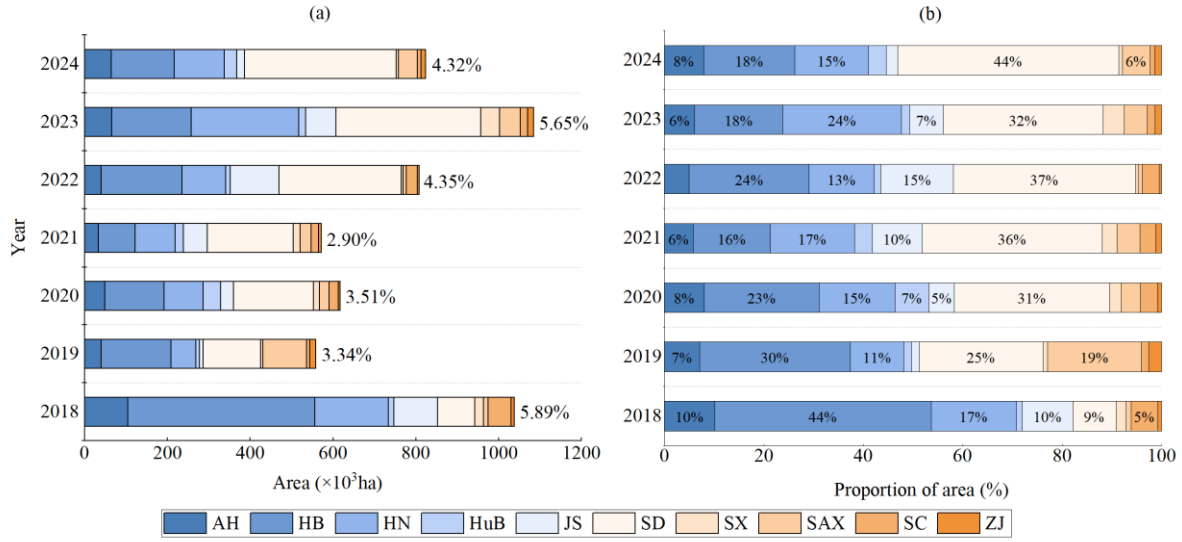

Figure R5: Differences between planted and harvested wheat area by province from 2018 to 2024. (a) Annual difference between planted and harvested wheat areas (in hectares) and its proportion of the total planted area; (b) Annual percentage of planted–harvested area difference for each province.

Point 4. The authors explained in the response the differences between the number of available provinces for the harvested area maps and the planted maps. This point is worth mentioning in the manuscript as well.

Response: Thank you for your helpful suggestion. We agree that the difference in the number of provinces available for the harvested area maps and the planted area maps is an important point that should be clarified in the manuscript. In the revised version, we have added an explicit explanation to the study area description section. The new text states that the ten provinces in eastern and southern China have large and contiguous winter wheat cultivation areas, which allows both planted and harvested area maps to be generated, whereas the five provinces in northern and northwestern China have limited and fragmented wheat cultivation, with mixed spring and winter wheat planting and similar harvest periods, so only harvested area maps were produced for these regions. This addition ensures that the rationale is clearly communicated in the main text.

P5, L135-141: Among the 15 provinces included in this study, the ten provinces located in eastern and southern China, which encompass the Huang–Huai–Hai Plain and the middle and lower reaches of the Yangtze River Plain, constitute the country's core winter wheat production regions. These areas are characterized by large cultivation scales and highly contiguous fields, which enables the extraction of both planted area maps and harvested area maps. In contrast, in the five provinces located in northern and northwestern China, wheat cultivation is relatively limited and fragmented, and some regions involve mixed planting of spring wheat and winter wheat with harvest periods that occur close to each other. As a result, only harvested area maps were generated for these provinces in this study.

Point 5. Figure 1(a) and Figure 6 display agro-agricultural zoning, this may lead to confusion to the fact that the paper actually follow province-level zoning

Response: Thank you for pointing this out. To avoid any potential confusion, we have removed the agro-agricultural zoning from the figures and replaced it with the corresponding province-level divisions. The revised content is now presented in the updated Fig. R6- R7.

The corresponding content has been made on Line 145 in Page 5 of the revised manuscript.

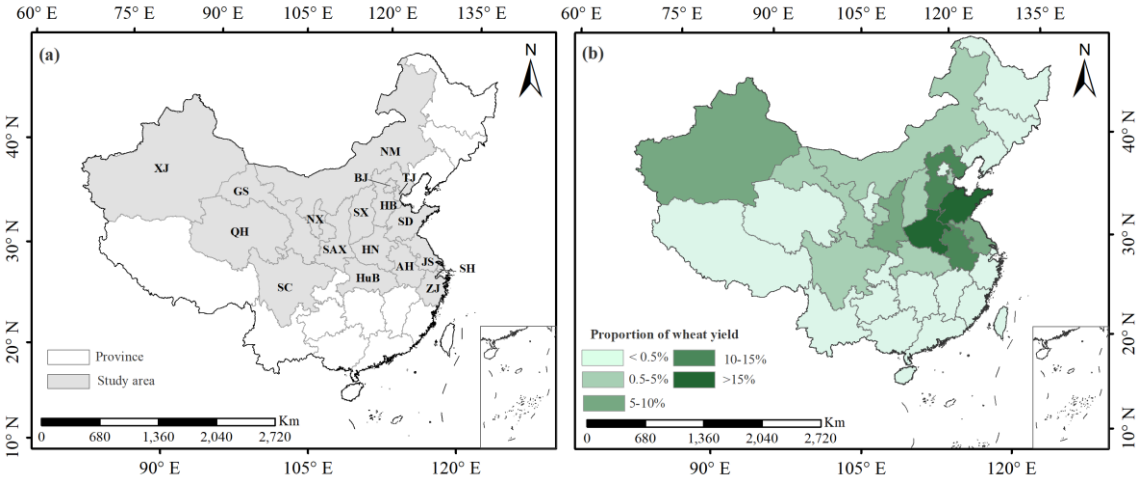

Figure R6: Location of study area in China. (a) Location of 15 provinces and 3 municipalities included in the study area. (b) Proportion of wheat production in 2022.

The corresponding content has been made in Page 13 of the revised manuscript.

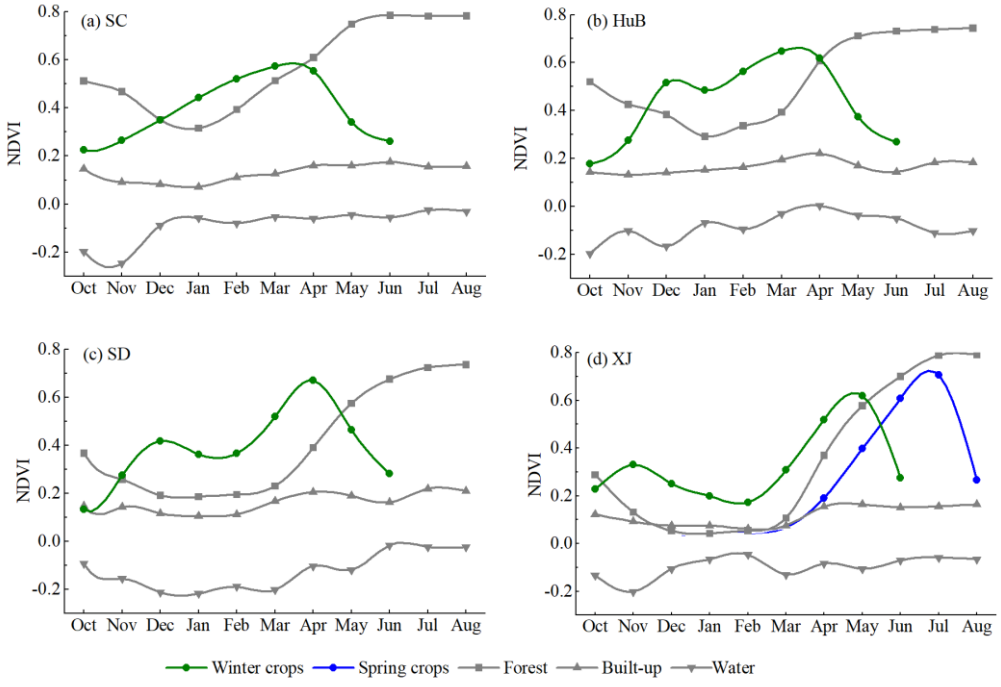

Figure R7: NDVI curves for different land cover types in four provinces.

Point 6. Please clarify how Figure 4 was generated. Confirm whether the samples used for VH analysis are independent from those used for validation.

Response: We thank the reviewer for raising these points, which we agree are crucial for clarity. Please find our clarifications below:

Generation of Figure 4: This figure was created to analyze the distribution of different land cover types within the derived wheat probability map. For each class, we randomly selected 500 sample points from the field survey data and plotted their frequency distribution against the wheat probability values.

P10, L241-243: To analyze the distribution of different land-cover types within the derived wheat probability map, we randomly selected 500 sample points for each land-cover category from the field survey data and plotted their frequency distribution against the corresponding wheat probability values.

Independence of Samples: We confirm that the samples for the VH analysis were indeed a randomly selected subset of the broader validation samples. This approach was taken to ensure the representativeness of the analysis while using our existing high-quality ground truth data. We acknowledge that they are not an independent set, and this point is now explicitly stated in the revised text to prevent any potential misunderstanding.

P10, L246-248: It should be noted that the sample set used for this VH analysis constitutes a randomly selected subset of the overall validation samples. This approach ensures the representativeness of the analysis while using the existing high-quality ground-truth data.

Point 7. Figure 19 requires more explanation in the caption (e.g., acquisition time, types of misclassified land cover).

Response: Thank you for the constructive suggestion. We have expanded the caption of Fig.R8 to include additional details such as the acquisition time of the imagery and the non-wheat area that were misclassified.

The corresponding content has been made in Page 30 of the revised manuscript.

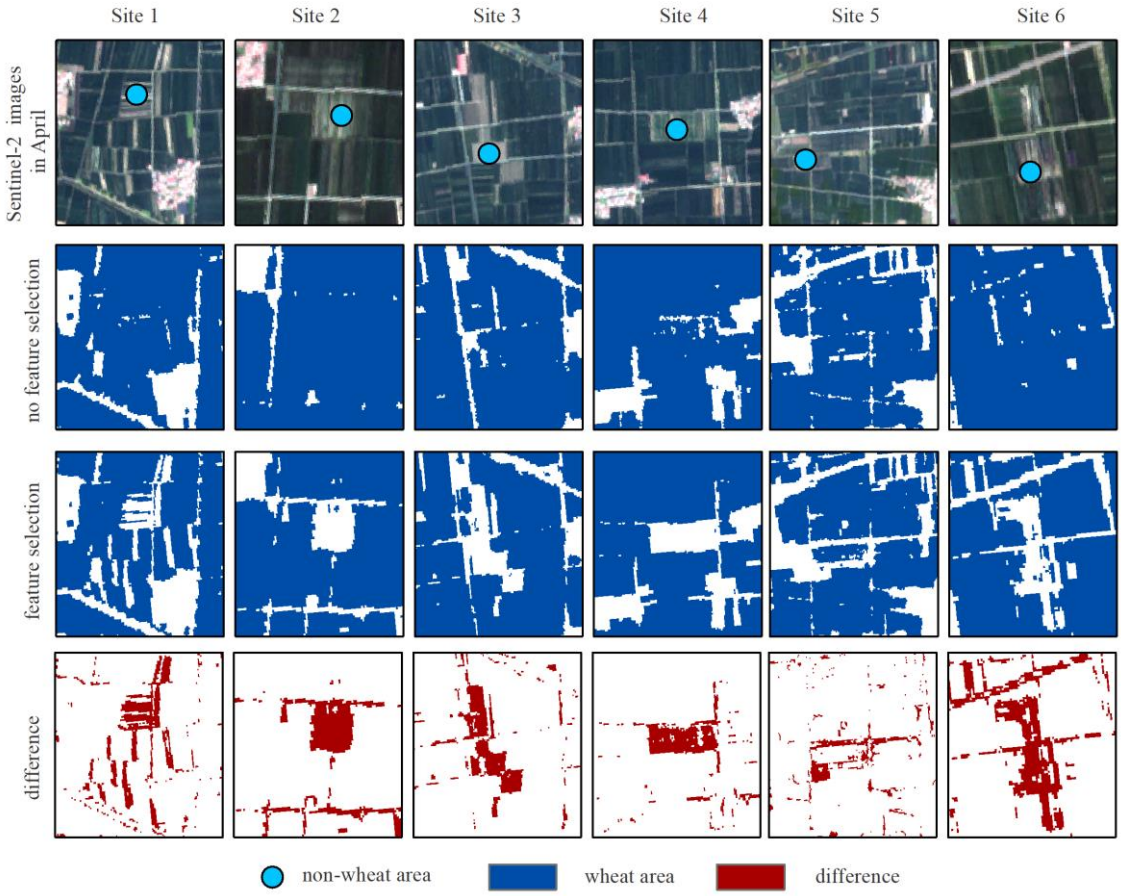

Figure R8: Comparison of wheat remote sensing recognition regions before and after feature selection.

Point 8. Line 225: The detailed explanation on how Figure 3 was generated could be moved to its caption to make the text more concise.

Response: We thank the reviewer for this valuable suggestion. We have moved the detailed explanation regarding the generation of the Figure from the manuscript to the figure caption to enhance the conciseness of the text. The revised caption now fully describes the data sources and methodology, as shown below.

P9, L219-228: Kansas and North Dakota are representative of winter and spring wheat systems, respectively, and their mid-latitude conditions result in strong phenological alignment with China's major wheat zones. As illustrated in Fig. R9, the phenological profiles from the United States closely match those of the corresponding wheat types in China, confirming the representativeness of the constructed samples.

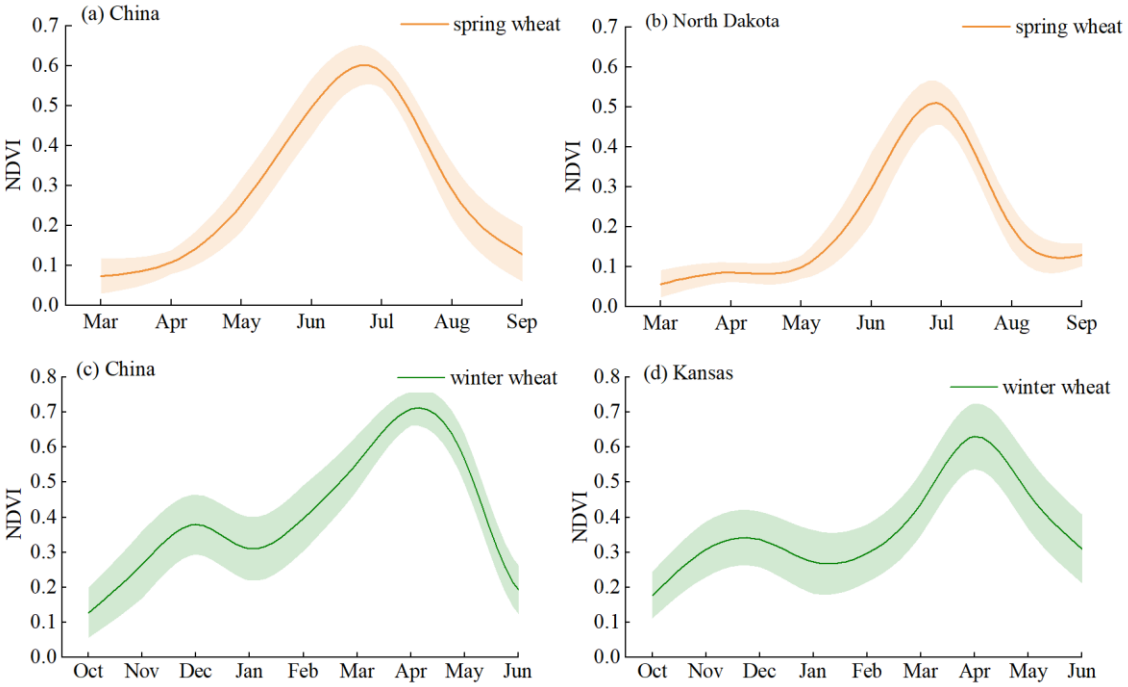

Figure R9: Comparison of NDVI time series curves between spring and winter wheat in China and the United States. (a) Spring wheat NDVI profile from field survey data in Northwest China. (b) Spring wheat NDVI profile from CDL data in North Dakota, USA. (c) Winter wheat NDVI profile from field survey data in the eastern plains of China. (d) Winter wheat NDVI profile from CDL data in Kansas, USA.

Point 9. Please explain how the number of samples for training and validation was determined, whether it is supported by any reference, and whether the sampling design is reasonable.

Tyukavina, A., Stehman, S. V., Pickens, A. H., Potapov, P., & Hansen, M. C. (2025). Practical global sampling methods for estimating area and map accuracy of land cover and change. Remote Sensing of Environment, 324, 114714. https://doi.org/10.1016/j.rse.2025.114714.

Response: Thank you for your valuable comment. Your question regarding the basis for determining the numbers of training and validation samples, as well as the overall reasonableness of the sampling design, is essential for ensuring the rigor of our study. Below we provide a detailed explanation of the procedures used to determine the training and validation sample sizes

Determination of the training sample size: The number of training samples was established using a data-driven and empirically based accuracy convergence analysis. Specifically, within each 0.5° grid cell, we progressively increased the number of randomly generated sample points from 20 to 1000 and systematically evaluated how classification accuracy changed with increasing sample size. The results indicated that when the numbers of wheat and non-wheat samples within each grid reached approximately 500 each, the model accuracy stabilized and no longer showed substantial improvement with additional samples. Therefore, we selected 500 wheat samples and 500 non-wheat samples as the most appropriate training sample size.

This choice ensures strong model performance while avoiding unnecessary computational costs. The relevant description is provided in the original manuscript on page 10, lines 257 to 261, as well as in the supplementary materials.

P10, L257-261: To ensure both regional representativeness and class balance, the number of samples in each province was determined based on a standardized grid approach, whereby each 0.5° × 0.5° grid cell was required to contain 500 sample points for wheat and 500 for non-wheat. This design effectively supports the robustness and generalizability of the classification model across heterogeneous agro-ecological zones. The sample size selection process is shown in Fig. R10.

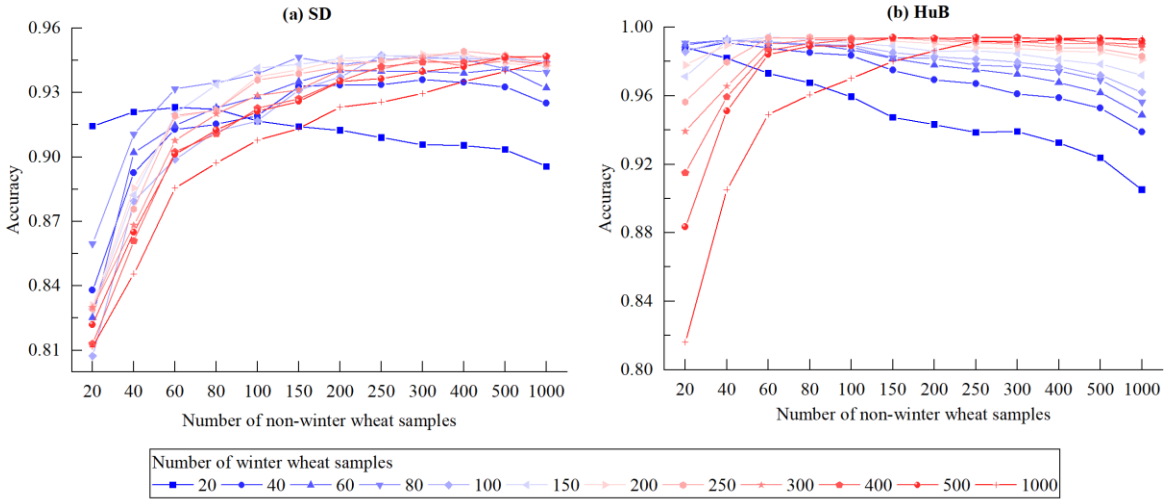

Figure R10. Changes in accuracy of winter wheat mapping under different sample sizes. (a)Variation trend of wheat accuracy in Shandong (SD) province. (b)Variation trend of wheat accuracy in Hubei (HuB) province.

Sampling design for validation samples: The sampling design for validation samples was developed with reference to methods used in previous studies (Liu et al., 2024; Liu and Zhang, 2023) and was adjusted to suit the specific characteristics of our research area. This approach ensures both scientific rigor and methodological appropriateness. We first divided the study area into regular quadrilateral grids, which served as the basic spatial units for sampling. All sample points were then manually interpreted using monthly Sentinel-2 composite images from the growing season, and were assigned to one of six land-cover categories: wheat, built-up land, water, trees, bare land, and other crops. For sample allocation, we applied a combined strategy that incorporates proportional allocation and category balancing. In grids containing wheat, we randomly selected between one and eight wheat samples according to the proportional wheat coverage within each grid, and supplemented these with non-wheat samples to maintain category balance. In grids without wheat, we randomly selected one to two non-wheat samples for each of the four categories, namely built-up land, trees, water, and bare land. The total number of samples per grid was limited to no more than ten, which effectively reduces excessive spatial clustering and minimizes the influence of spatial autocorrelation on the validation results.

This sampling strategy ensures a spatially uniform distribution of samples and provides representative coverage of diverse land-cover types, while also adhering to the fundamental principles of stratified sampling in statistics. A detailed description of this method is provided in the manuscript on page 6, lines 175 to 179, as well as in the supplementary materials. To further enhance the rigor of the study, the corresponding literature references have been added in the revised version.

P6, L175-181: The sampling design of validation samples refers to the methods in previous studies (Liu et al., 2024b; Liu and Zhang, 2023), and has been adjusted according to the specific conditions of this study to

ensure its scientific and rationality. Multi-temporal Sentinel-2 imagery from 2017 to 2024 was dynamically explored through the Google Earth Engine (GEE) visualization platform. Manual interpretation was conducted by combining spectral, textural, and temporal variation characteristics. A spatially stratified sampling strategy based on quadrilateral grids was adopted to mitigate the effects of spatial autocorrelation. To further improve interpretation accuracy and boundary delineation, historical very high-resolution imagery (GE-VHR) from Google Earth was used for auxiliary verification.

Liu, W., Li, S., Tao, J. et al. CARM30: China annual rapeseed maps at 30 m spatial resolution from 2000 to 2022 using multi-source data. Scientific Data 11.1 (2024): 356. https://doi.org/10.1038/s41597-024-03188-1.

Liu, W. and Zhang, H. Mapping annual 10 m rapeseed extent using multisource data in the Yangtze River Economic Belt of China (2017–2021) on Google Earth Engine. International Journal of Applied Earth Observation and Geoinformation 117 (2023): 103198. https://doi.org/10.1016/j.jag.2023.103198.

P3-4 in supplementary data: As shown in Fig.R11, we take Xinyang City in Henan Province as an example to demonstrate the spatial hierarchical sampling strategy based on quadrilateral grids. We first divided the study area into a series of regular quadrilateral grids within the Google Earth Engine platform and examined the monthly Sentinel-2 composite imagery from October to the following June. During the interpretation process, all land cover types were classified into six categories: winter wheat, built-up, water, trees, bare land, and other crops. Based on the proportional coverage of each land cover type within a grid, no more than 10 sample points were randomly selected per grid, thereby avoiding excessive spatial clustering of samples. Specifically, for grids containing wheat fields, 1–8 wheat samples were randomly selected in proportion to the wheat coverage within the grid, along with additional non-wheat samples to maintain category balance. For grids without wheat fields, 1–2 non-wheat samples were randomly selected for each category, including built-up land, trees, water bodies, and bare land. This approach ensured both the spatially uniform distribution of samples and the representativeness of diverse land cover types.

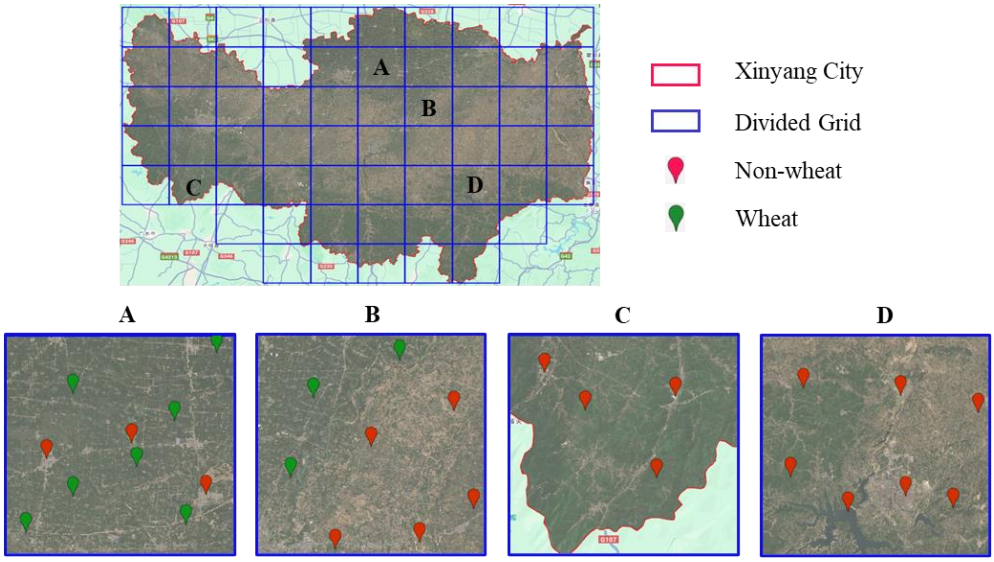

Figure R11: Spatial stratified sampling strategy in Xinyang City.