https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Open Access Earth System Science Data Discussions

Mapping 20-years winter wheat dynamics in global primary planting areas using Gaussian mixture models with adaptive thresholds

Yanan Wen[1,2#], Tuo Chen[3#], Xuecao Li[3*], Tiecheng Bai[1*], Ke Yao[4], Liheng Zhong[5], Han Chen[6], Meiling Liu[2], Xieqin Huang[7,8], Shunlin Liang[9], Shuangxi Miao[4], Jianxi Huang[4,10*]

[1] Key Laboratory of Tarim Oasis Agriculture (Tarim University), Ministry of Education, Alar, 843300, China
[2] School of Information Engineering, China University of Geoscience, Beijing, 100083, China
[3] Guangdong Key Laboratory for Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-Sen University, Guangzhou, 510006, China
[4] College of Land Science and Technology, China Agricultural University, Beijing, 100083, China
[5] Ant Group, World Financial Center, Beijing, 100000, China
[6] Swiss Reinsurance Company Ltd Beijing Branch, Beijing, 100022, China
[7] Department of Computer Science and Technology, Tsinghua University, Beijing, 100190, China
[8] School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou, 510275, China
[9] Department of Geography, The University of Hong Kong, 999077, Hong Kong, China
[10] Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, 611756, China

# These authors contributed equally.

Correspondence to: Xuecao Li (lixc26@mail.sysu.edu.cn); Tiecheng Bai (baitiecheng@taru.edu.cn); Jianxi Huang (jxhuang@cau.edu.cn)

Abstract. Understanding the spatiotemporal dynamics of winter wheat is essential for ensuring global food security. Currently, limited research has focused on the global dynamics of wheat over past decades. In this study, we propose a novel framework to map fractional winter wheat dynamics from 2001 to 2020 at 1 km resolution in key global planting areas from MODIS satellite data, utilizing a flexible Gaussian mixture model. We first created the stratified samples of winter wheat fractions at 1km resolution from multiple public crop datasets, and then developed a robust random forest regression model using MODIS surface reflectance. Subsequently, we estimated the actual wheat cover fractions across different regions and years by analyzing crop mixtures within 1°×1° grids with multiple Gaussian models. The model parameters were utilized to determine optimal thresholds for winter wheat extraction. The performance of our proposed framework was evaluated spatially and temporally, revealing significant insights into global winter wheat dynamics. Results demonstrated that our mapping approach aligns closely with existing local winter wheat maps and statistical data, achieving a coefficient of determination (R^2) of 0.81 with FAO statistics in primary planting regions and exceeding 0.72 at subnational scales. This study presents the first comprehensive effort to map global winter wheat distribution and dynamics from 2001 to 2020 at a near-global scale. The proposed framework is readily adaptable to other major crops and demonstrates strong agreement with existing maps and statistical records. The resulting high-resolution global winter wheat map series provides valuable inputs for global crop modeling and contributes to achieving the "Zero Hunger". The product is publicly available at https://doi.org/10.6084/m9.figshare.32149033.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Open Access Earth System Science Data Discussions

1. Introduction

As one of the world's most important cereal crops, wheat plays a crucial role in global food security and the sustainability of agricultural systems. Ongoing population growth and the increasing frequency of extreme weather events are intensifying pressures on cropland systems, posing substantial challenges to global food production (Blickensdörfer et al., 2022; Yang et al., 2023). Among wheat types, winter wheat exerts a preponderant influence on the global production landscape, encompassing a substantial proportion of the total harvested wheat area, estimated to exceed three-quarters (Jiang et al., 2022; Wójcik-Gront, 2018). For example, winter wheat contributes approximately 70–80% of total wheat production in the United States (Ren et al., 2019) and more than 95% in China and Ukraine (Júnior et al., 2022; Zhang et al., 2022b). Globally, winter wheat cultivation exhibits strong spatial heterogeneity and long-term variability due to differences in climatic conditions, farm size, and field fragmentation (Zhang et al., 2019). Mapping the long-term dynamics of winter wheat at a global scale for agricultural activities (e.g., early warning of global food crises, agricultural land optimization, and sustainable decision-making) is an extremely challenging yet critically important task (Hunt et al., 2019; Nelson and Burchfield, 2021).

The development of new sensors and satellite platforms has greatly improved the efficiency and accuracy of crop mapping using remote sensing techniques (Ashourloo et al., 2022). However, efficiently compositing time-series observations to extract key phenological information during crop growth remains challenging because of frequent cloud contamination and the relatively long revisit intervals of sensors such as Landsat (Zhu et al., 2022). Despite the relatively short revisit cycle for Sentinel-2 data (i.e., 5 days), its relatively short period and spatial resolution variation, making it difficult to widely apply in long-term crop mapping (Kimambo and Radeloff, 2023). Moreover, these optical data also pose limitations regarding data storage (Kempeneers and Soille, 2017). Over past decades, the Moderate Resolution Imaging Spectroradiometer (MODIS) data has been widely applied for crop mapping and its dynamics at a large scale (Zhong et al., 2019). Notably, using MODIS images with high temporal availability can depict the spectral dynamics of crop types during the entire growth period to retrieve subtle but crucial phenological information in the different growing stages (Huang et al., 2022; Qu et al., 2021).

Only limited efforts have been made to map crop dynamics at large spatial scales, particularly using coarse-resolution satellite observations (e.g., 250 m, 500 m, and 1000 m). This is because the disturbance of mixed pixels and spectral confusion within crop regions may affect the mapping performance (Zhang et al., 2021). To mitigate such a problem, model-driven approaches (e.g., linear and nonlinear models) and data-driven approaches are widely employed through decomposing crop spectral signals (Wang et al., 2015); as such, fractional information is more reliable than binary pixels, especially at relatively coarse resolution. Although several fractional mapping models (e.g., linear spectral mixture models, multiple endmember-based approaches, and regression methods based on machine learning technology) have been developed with relatively good performance at regional scales (Ghaffari et al., 2017; Kathirvelu et al., 2023), they are inapplicable to support large-scale applications, due to spectral variability within and among different crop types. Fortunately, the Gaussian Mixture Model (GMM) offers a probabilistic interpretation with sufficiently smooth identification boundaries, enabling reliable

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

mapping reliable crop dynamics (Wang et al., 2019). Moreover, the GMM demonstrates excellent spectrum unmixing capabilities for coarse resolution satellite observations (Chen et al., 2018). However, most studies about crop monitoring based on the GMM have been mostly limited to cluster and area estimation for different crop types (You et al., 2023), and few studies have been directly used to investigate large-scale crop dynamics across space and time (Harshvardhan et al., 2020). Therefore, using crop fraction and the GMM within each grid, we developed a robust, universal detection model for fractional mapping.

Although several global and national datasets of crop extent with temporal and spatial dynamics have been proposed (Dong et al., 2020; Shen et al., 2024), few studies have directly mapped crop fractions. Currently, crop type maps use only binary maps (e.g., wheat and non-wheat) to represent single crop distribution across time and space (Ghassemi et al., 2022). Most fractional maps of crops in agricultural practices are mainly aggregated from high resolution maps (e.g., 10m to 30m), which is difficult to implement at the global scale and across multiple years (Ma et al., 2019; Rufin et al., 2022), to support analyses such as agricultural status (e.g., cropland expansion, fallow, and crop distribution patterns). Therefore, crop fraction map is critical to comprehensively understand temporal and spatial differences within and among homogeneous crop pixels, and further to effectively reduce uncertainty of fragmented and complex agricultural production structures (Hu et al., 2021). In addition, crop fraction map can be set as input parameters to support global crop growth model (e.g., APSIM) to improve model simulation ability (Hao et al., 2021).

Here, we developed a framework using gaussian parameter regression strategy to generate large-scale and long-term winter wheat fraction maps at the global scale, at 1 km resolution and spanning from 2001 to 2020. We derived fractional wheat samples from existing knowledge, i.e., publicly available winter wheat datasets at relatively fine resolution. Subsequently, we utilized a gaussian parameter-based method for the estimation of the optimal threshold from the estimated fractional maps. Finally, we assessed the mapping performance from spatial and temporal perspectives using multiple datasets from available crop cover datasets and official statistics.

2. Study area and datasets

2.1. Study area

To capture global winter wheat dynamics, we selected 53 countries representing the main winter wheat cultivation regions (Fig.1). To collect sufficient samples of winter wheat proportion across different countries, we divided these countries into two parts (i.e., pink and blue regions in Fig.1) regarding available winter wheat datasets (e.g., the winter wheat classification datasets and yield data). Here, the region included in our analysis represents most of global wheat cultivation, encompassing about 92% of the harvested areas and 93% of global production (FAOSTAT, 2020).

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

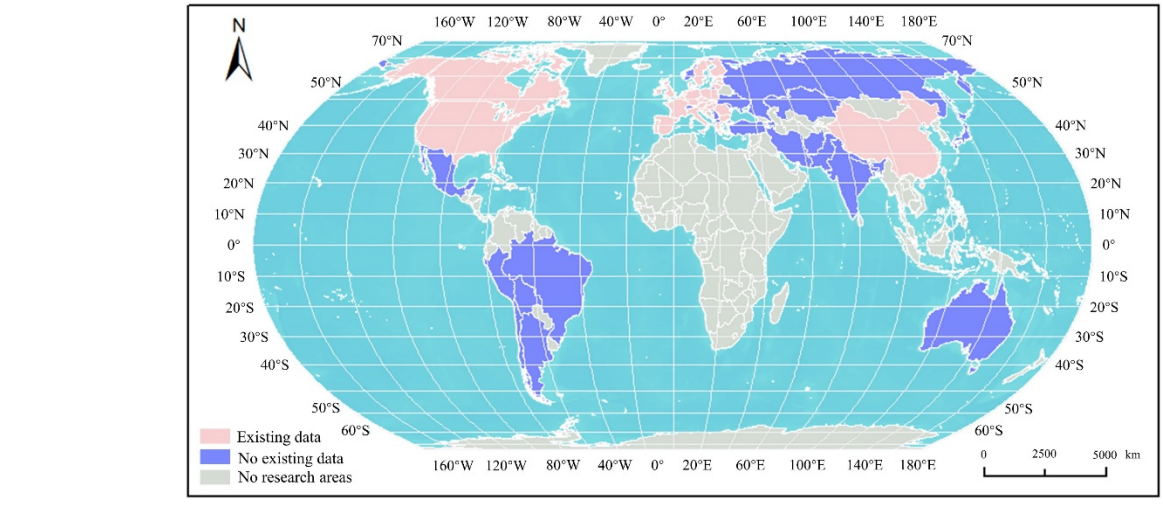


Fig 1. The primary planted areas of winter wheat in our study, including Asia, Europe, Oceanica, North America and south America.

2.2. Datasets

We adopted daily MOD09GA v061 images as our primary dataset, spanning from 2001 to 2020, to map the dynamics of winter wheat in a form of fraction. Specifically, we used all available MOD09GA v061 surface reflectance images to better capture the seasonal dynamics of winter wheat (Ding et al., 2022), with seven calibrated daily spectral bands with spatial resolutions of 500 km, of which the atmospheric effect has been corrected.

We used four crop type datasets to collect winter wheat samples with proportion information. The Cropland Data Layer (CDL) were produced by the National Agricultural Statistics Service (NASS) in the United States Department of Agriculture (USDA) at 30m resolution , which provides annual crop cover from 2008 with mean accuracies exceeding 95% (USDA-NASS, 2023). For Canada, the Agriculture and Agri-Food Canada (AAFC) provides an Annual Crop Inventory (ACI) product using high-resolution remotely sensed imagery, to depict national crop cover at a 30 m resolution from 2011 to present (AAFC, 2023). The EUCROPMAP product offers detailed spatial information of 19 crop types across Europe (EU) in 2018 with 10 m resolution and high accuracy (Shao et al., 2023). In addition, Dong et al. (2020) generated the China winter wheat maps at 30 m resolution with high accuracy. To characterize the winter wheat fractional changes at the pixel scale (i.e., 1km), we aggregated 10m and 30m binary winter wheat maps into 1km winter wheat fractional maps. We utilized all acquired winter wheat proportion samples to construct the regression model for estimating their wheat proportions. Besides, they were also used to derive locally optimal thresholds using parameters from multiple gaussian models.

In addition, we also collected global winter wheat yield data (i.e., GlobalWheatYield4km) as a reference for countries where winter wheat datasets were not available (Zhang et al., 2024). The GlobalWheatYield4km dataset provides pixel-level yield

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Open Access Earth System Science Data Discussions

information of spring and winter wheat at a 4km resolution. We aggregated 4km binary winter wheat maps into the 10km fractional maps to collect proportion samples. Given that the GlobalWheatYield4km data covers the whole period over the past two decades, here we temporally expand extracted winter wheat samples from 2020 to other years. Moreover, we collected cultivated acreage statistics data from the Food and Agriculture Organization (FAO) to validate our mapping results. Specifically, we utilized the crop acreage of wheat from the Crops and Livestock Products at the country level to validate our temporal and spatial results from 2001 to 2020 (FAOSTAT, 2020).

3. Methodology

This study presents a framework for capturing the long-term and large-scale dynamics of winter wheat across key producing regions worldwide (Fig.2). First, using collected winter wheat samples in each country, we estimated the fraction of winter wheat within each 1 km resolution over past decades (Fig.2a). These winter wheat samples derived from proportion maps aggregated by crop type map, using the random stratified sampling approach. Then, we identified the winter wheat dynamics with estimated locally optimal thresholds, using multiple gaussian curve parameters across time and space (Fig.2b). Specifically, we extracted gaussian parameters by decomposing gaussian curves, and these parameters were served as inputs to estimate the locally optimal thresholds within each 100 km grid. Finally, we assessed the mapping performance across spatial and temporal perspectives using multiple datasets from various sources (i.e., publicly available crop datasets and official statistics) (Fig.2c).

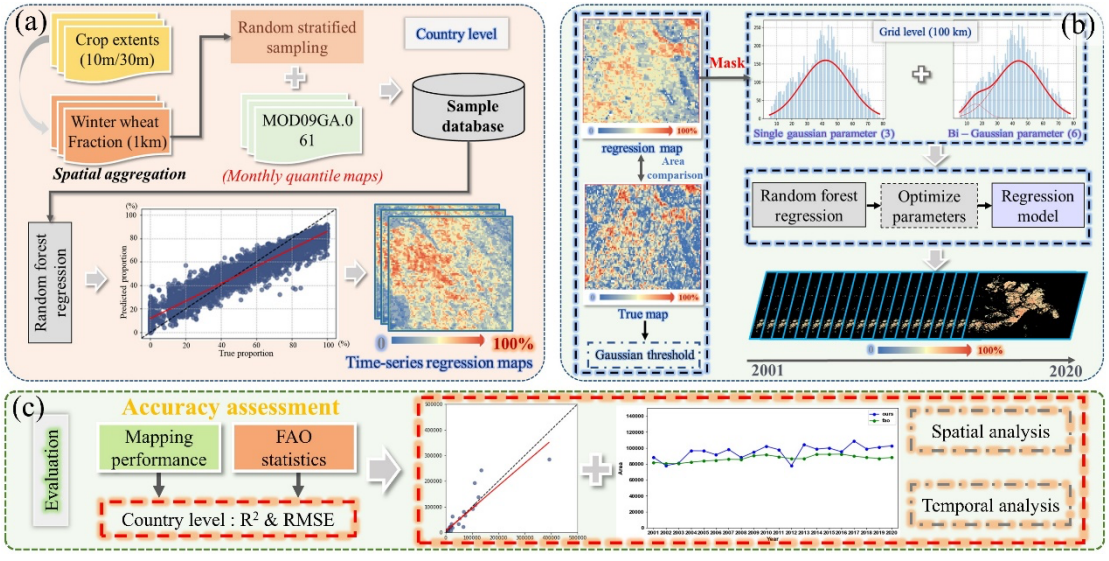


Fig 2. The methodology of winter wheat mapping with gaussian parameter regression strategy, including the generation of regression proportion map from the winter wheat products (a), the estimation of locally optimal thresholds using gaussian parameters (b), and the temporal and spatial evaluation of mapping performance (c).

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

3.1. Generating regressed fraction maps of winter wheat

We generated fractional maps of winter wheat using the robust and harmonious random forest regression model in each country (Fig.3). Due to their high spatial resolution or long-term data availability, several crop datasets, i.e., CDL (Borrelli et al., 2023), EUCROPMAP (Hao et al., 2020), China winter wheat maps (Dong et al., 2020), and GlobalWheatYield4km (Zhao et al., 2023), have been widely utilized in agricultural fields. Thereafter, we adopted these data to extract reliable training samples for winter wheat with true fraction references using a random stratified sampling algorithm. In our study area, we first aggregated binary maps with fine spatial resolution (i.e., 10m or 30m) of winter wheat into 1 km using the mean filter to generate the fractional map, which represents the proportion of winter wheat within each pixel (Fig.3a). Particularly, in regions where crop data are unavailable (e.g., India, Australia, and Brazil), we aggregated the 4 km GlobalWheatYield4km dataset to 10 km and incorporated as ancillary data for deriving fractional samples.. Then, we used random stratified sampling scheme to collect training samples for winter wheat proportion estimation in each country (Fig.3b). Specifically, we divided each regressive map from 0 to 100% using the interval of 10%, based on which samples were randomly extracted within each category. Finally, we mapped the winter wheat dynamic in a fractional form from 2001 to 2020 by introducing the previously developed “adaptive strategies” approach (Fig.3c) (Wen et al., 2022), which aims to use representative samples with repeated sampling to improve the model performance. It is worth noting that this is an iterative process of training the model for robust performance.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

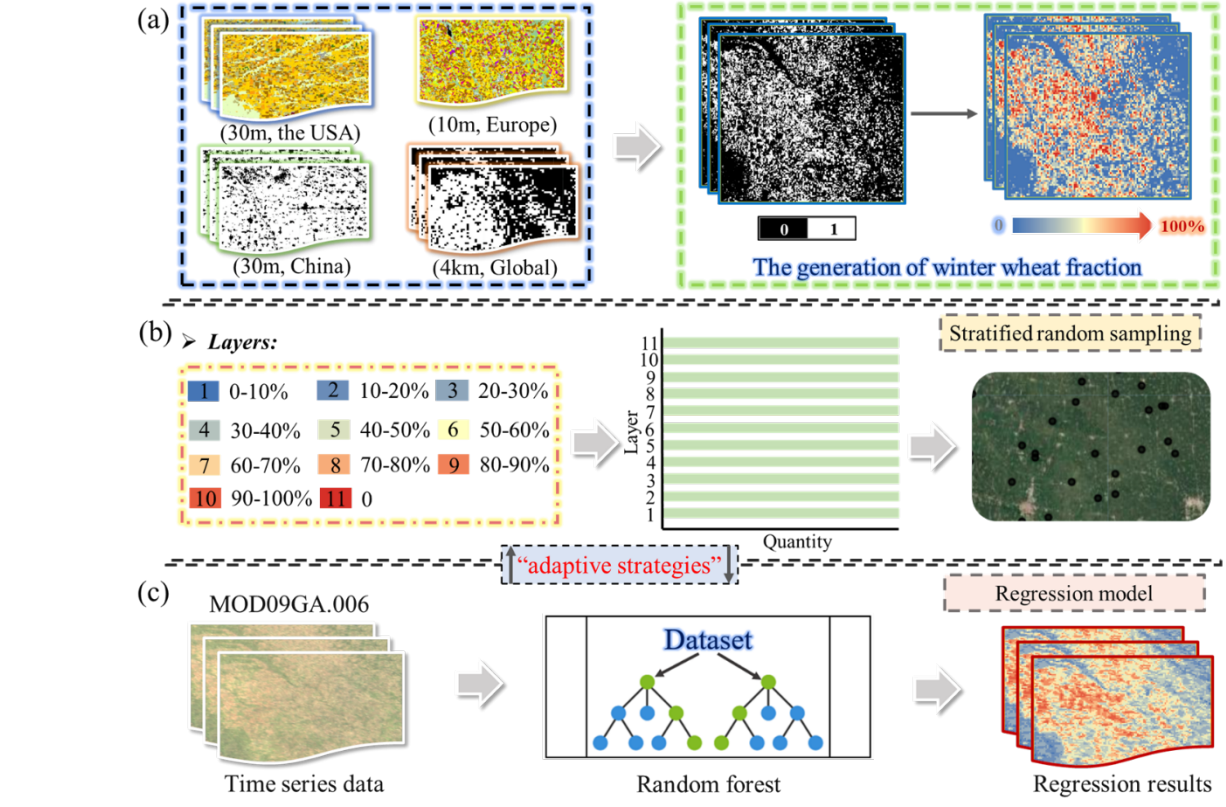
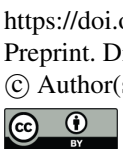


Fig 3. Mapping the winter wheat fraction using samples from existing datasets, including the generation of winter wheat fraction maps (a), stratified random sampling (b), and establishment of winter wheat regression model (c).

3.2. Predicting the locally optimal threshold using gaussian mixture parameters

We analyzed the gaussian parameters within the 100 km (i.e., 1°) grids to train the global random forest regression model regarding segmentation threshold (Fig.4). First, we established a global seamless grid (i.e., 100km×100km) for all countries in our study area, to extract gaussian parameters in the next step. Besides, we retained the cropland extent using the GLAD dataset and masked the non-cropland areas to reduce uncertainties (Potapov et al., 2022). In particular, we used the receiver operating characteristic (ROC) curve to derive the optimal threshold (balanced point) between referenced and predicted proportion maps, which is necessary to further exclude uncertain winter wheat pixels, such as those less than 0.206 in Fig.4a. Then, we analyzed gaussian parameters in both the single and bi-gaussian curves within each 100km grid as the training features, respectively (Fig.4b). The gaussian model shows clear physical meaning regarding the dominant crop types within (Guan et al., 2023). Taking the bi-Gaussian curve as an example, clear differences between the two distributions of winter wheat are observed, demonstrating the discriminative power of Gaussian parameters, including the mean (μ), standard deviation(σ), and amplitude(A) (Fig. 4b). Considering the possibility of relatively high proportion within the grids in the

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026



major planted regions of winter wheat, we also adopted the parameters derived from single gaussian function to enrich the diversity of input features. Finally, we estimated the locally optimal threshold for winter wheat proportion using these gaussian features as inputs for each grid, under which value areas are relatively consistent between referenced and predicted fractional maps (Fig.4c). We further divided the collected samples into training (70%) and validation (30%) datasets here. That is, we developed the robust regression model of winter wheat identification based on training samples to estimate the optimal thresholds for each local grids, while the remaining samples were used for validating threshold estimation in our study.

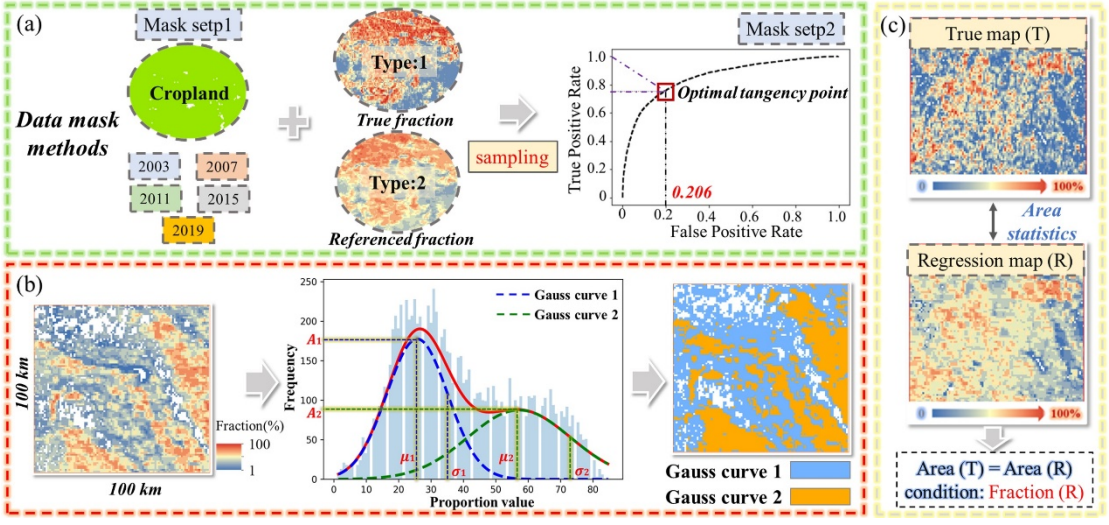


Fig 4. The estimation model of locally optimal threshold using gaussian mixture parameters, including two data mask steps (a), an illustration of the gaussian parameter for each curve within each grid (b), and the optimal threshold estimation corresponding to gaussian parameter within each grid (c).

3.3. Evaluating the mapping performance

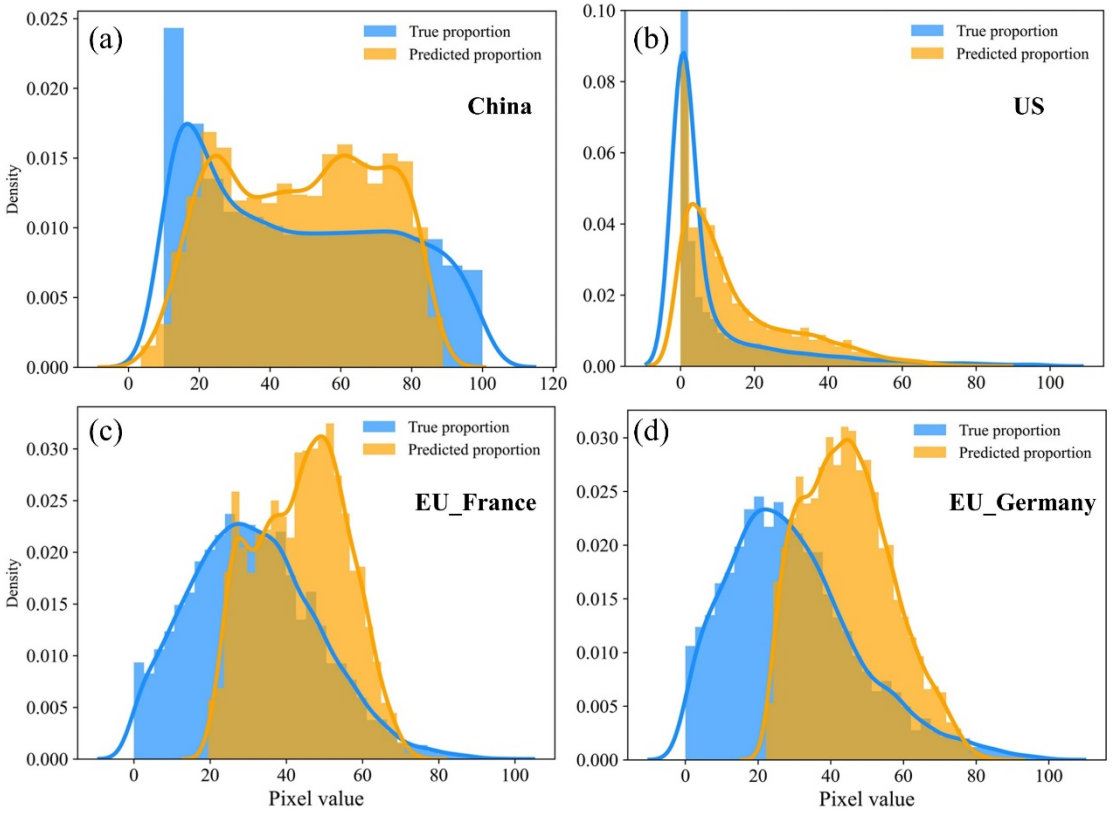
We evaluated our fractional mapping results at both the grid and country scales. At the grid scale, we used publicly available winter wheat datasets as reference data, and compared with aggregated winter wheat areas within 100 km grids to assess the performance of the Gaussian parameter-based threshold estimation approach. At the country scale, we employed the Pearson correlation coefficient (R), coefficient of determination (R^2), and root-mean-square error (RMSE) to evaluate the mapping performance of the derived fractional winter wheat results (Pott et al., 2021; Tran et al., 2022). We further compared winter wheat areas and their annual trends with surveyed records from the FAO database at country level (https://www.fao.org/faostat/en/#data). In addition, we also assessed the mapping area at a subregional scale between our mapping results and official statistics in several regions (e.g., China, the US, and EU countries).

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

4. Results and discussion

4.1. Winter wheat regression model

The spatial distribution between aggregated fraction maps from fine resolution (i.e., true proportion) and the regressive maps (i.e., predicted proportion) in winter wheat are relatively consistent, as illustrated in four countries (i.e., China, US, France, and Germany) (Fig.5). We visualized the distribution of 5000 cropland samples with fraction values in the form of frequency histogram in each country, which were derived from aggregated and regressive maps of winter wheat using randomly sampling approach, respectively. We found there is a relatively high similarity between true and predicted proportions in both China and US (Fig. 5, a-b), while those in France and Germany show some fraction gaps in low-proportion distributions (i.e., 0-20%) (Fig. 5, c-d). This phenomenon is attributable to (1) the systematic error associated with machine learning–based regression, characterized by the underestimation of high fraction values and overestimation of low fraction values (Yahiaoui et al., 2021), and (2) uncertainties arising from data quality limitations and variations in MODIS image availability over previous decades (Ren et al., 2021). Specifically, winter wheat fractions are highest in China and lowest in the United States, with intermediate values observed in France and Germany, which is highly related to cropping pattern and crop management (Liu et al., 2020).



https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Open Access Earth System Science Data Discussions

205 **Fig 5. The comparison of spatial distribution between aggregated fraction maps and regressive maps for winter wheat in China (a), the US (b), France (c), and Germany (d).**

4.2. Gaussian parameter-derived optimal thresholds

We developed a robust and harmonized gaussian parameter derived optimal thresholds for winter wheat fraction map with good performance (Fig. 6). Using the same validation dataset illustrated in Section 3.2, we compared the area differences of
210 winter wheat between true and predicted proportions at 100 km grid scale, to evaluate the robustness of the model in various regions, such as China, US, and EU countries. Overall, our predicted result (i.e., blue box) agrees well with the reference proportion (i.e., yellow box)(Fig. 6), suggesting the final model can achieve optimal thresholds using gaussian derived parameters, within 100 km grids at a large scale. Besides, other features such as violin plot, mean value (i.e., red rhombuses), and median (i.e., black bold lines) also show good agreement for these two fraction maps in different countries. In addition,
215 the winter wheat area within each grid in China is commonly larger than the other two regions, indicating that there is a relatively large gap in cropping patterns in these three countries, which may be related to regionalization and varieties standardization of wheat production in China (Jin et al., 2023).

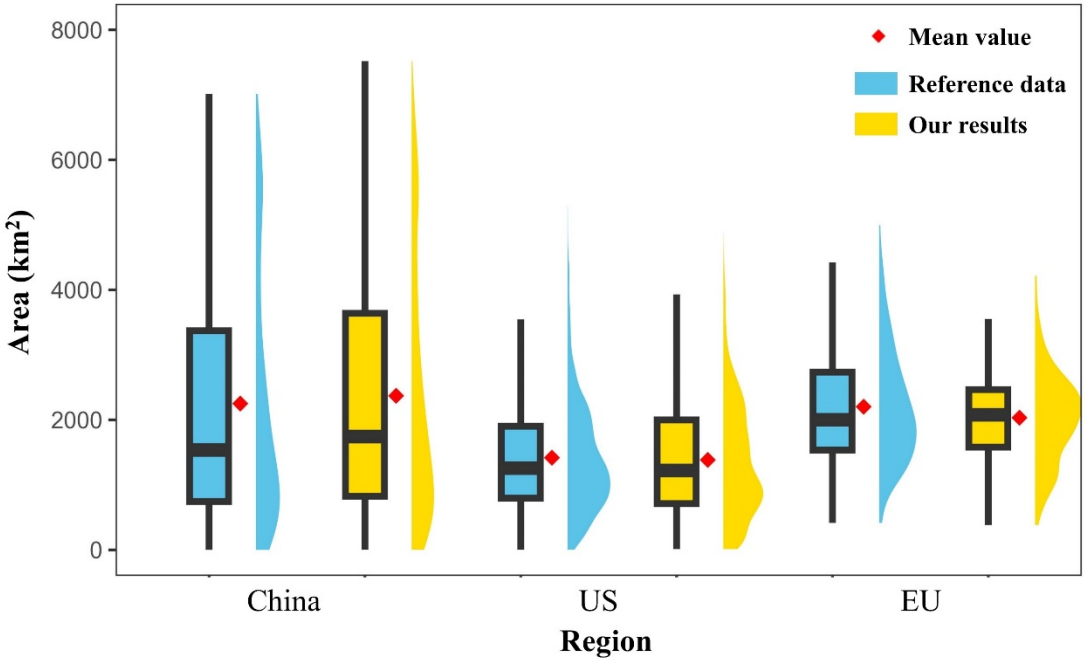
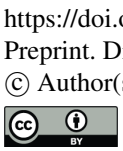


Fig 6. Comparison of the derived areas using gaussian parameter derived optimal thresholds in different countries or regions.

220 ### 4.3. Accuracy assessment of winter wheat fraction maps

4.3.1. Compared with country-level FAO statistics

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026



Open Access Earth System Science Data Discussions

The comparison of our mapping results with FAO acreage statistics from 2001 to 2020 indicates that winter wheat mapping accuracies are generally high across most countries in the study region (Fig. 7). The R^2 values, all above 0.8, confirm that the mapped winter wheat areas are in strong agreement with FAO statistics at the national scale. Besides, the slope between our fractional mapping results and statistics data tends to be 1, i.e., the differences in the winter wheat area of these two results are relatively small at the country scale. At the national level, our mapping results exhibit a similar annual trend to the reference data during 2001 to 2020, with the annual difference relative to the long-term mean within 10% (Fig.S1), suggesting our model can well capture the dynamics of winter wheat. Nevertheless, our mapped results of winter wheat fractions are slightly overestimated in most years (e.g., 2013 and 2019), which may be closely related to the systematic disturbance of the random forest regression model (Schwalbert et al., 2020) and the data quality (e.g., the available crop datasets and MODIS images) (Wang et al., 2020).

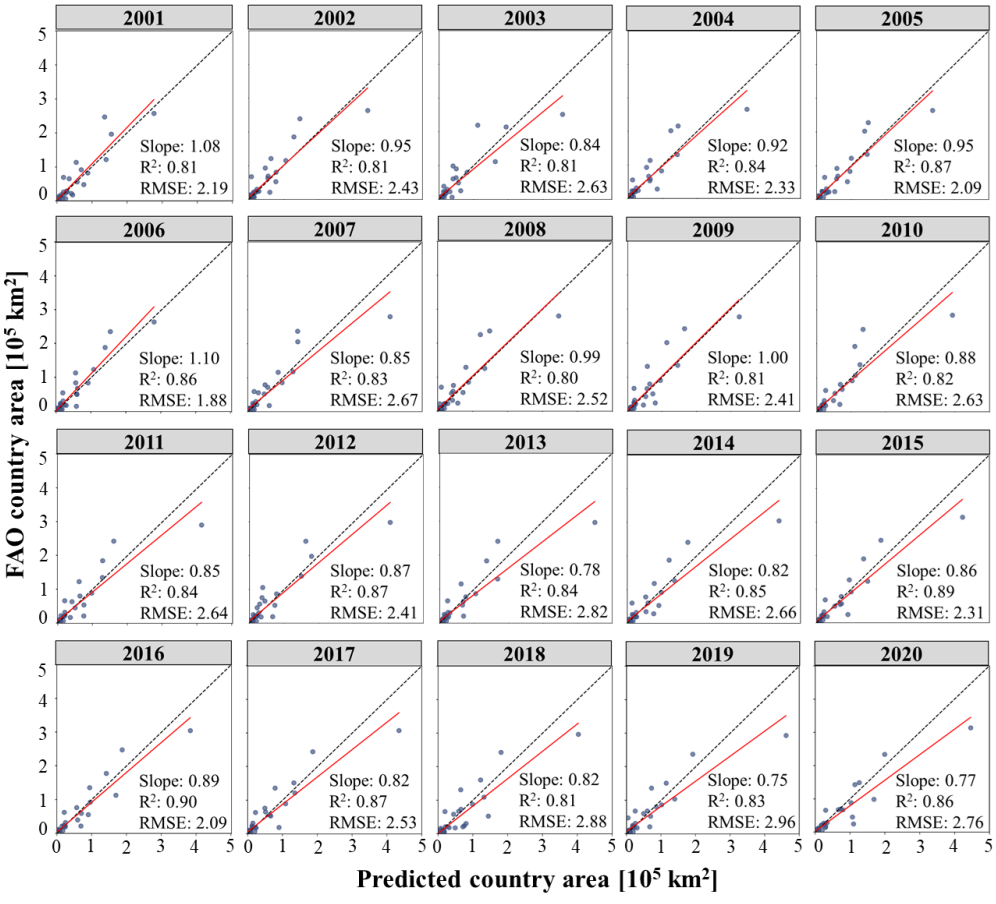


Fig 7. Comparison of national-level winter wheat areas derived from our mapping results and FAO statistical data over the period 2001–2020.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

235 **4.3.2. Compared the spatial pattern with other winter wheat products**

Comparing our mapped results in 2020 and the crop type datasets (i.e., CDL and the winter wheat product), most winter wheat growing regions (i.e., China and the US) of winter wheat have relatively similar spatial dynamics (Fig.8, 9). Here, we aggregated binary winter wheat maps at finer resolution into 1km winter wheat fractional maps, to integrate the comparison criteria to the fractional level. Our mapped results show strong agreement with the reference datasets, indicating that the
240 fractional mapping model achieves relatively high accuracy at a large scale. Enlarged views of three representative winter wheat growing regions are presented to further illustrate the performance of the gaussian mixture-based locally optimal threshold approach (Fig.8c, 9c). The derived fractional mapping results for sites A-C in China and the US are close to the acquired reference fraction data, although there exists a missing phenomenon in some distribution areas with a low fraction of winter wheat. In addition, our model could also achieve relatively good performance in EU winter wheat growing
245 countries (i.e., France, Poland, and Germany in 2018) (Fig. S2).

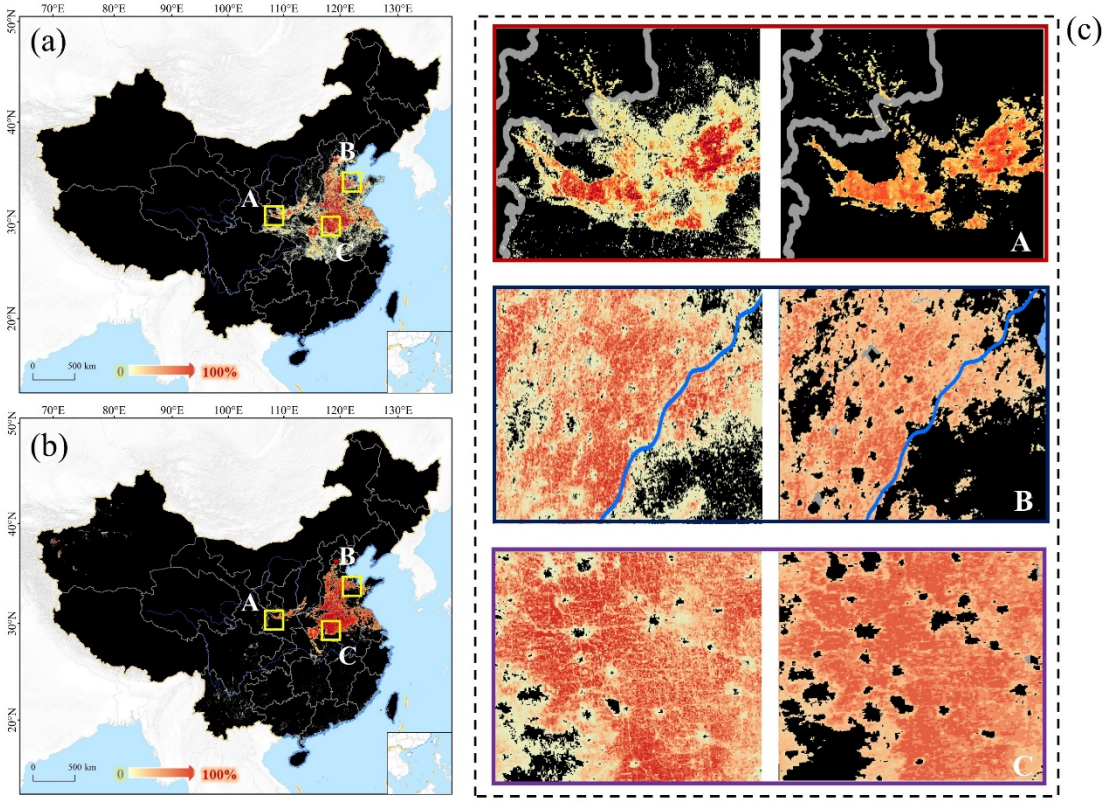


Fig 8. Comparison of mapping performance between the existing product (a) and our mapped results, (b) for China in 2020, (c) is enlarged views for typical regions. The left figure of (c) is the aggregated winter wheat map produced by Dong et al., (2020), and the right is our result.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

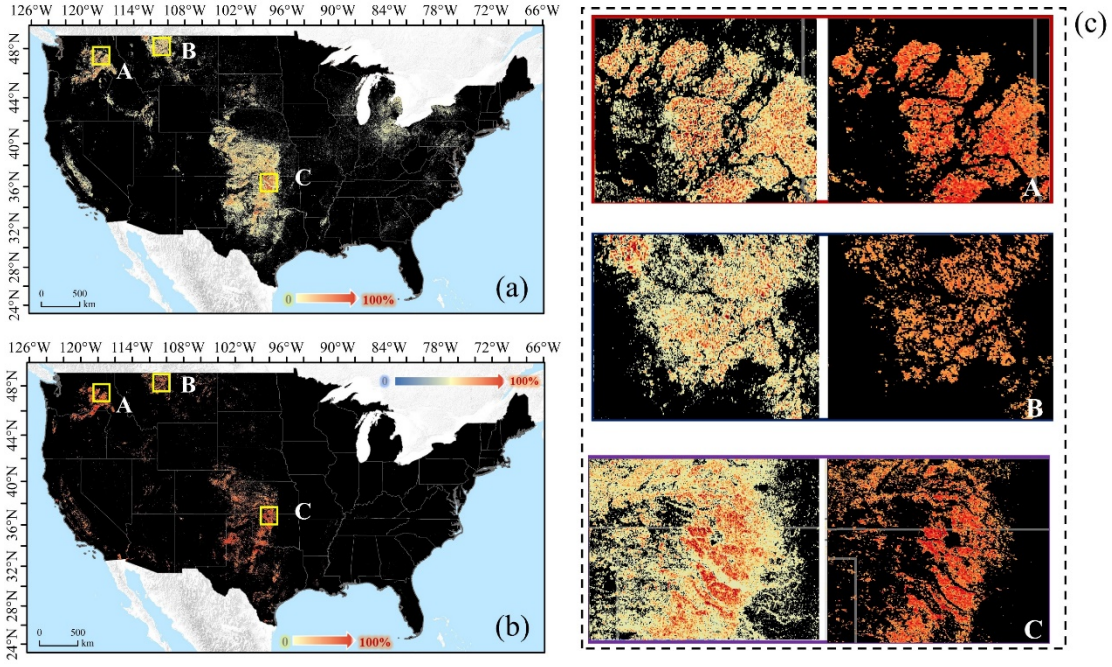


Fig 9. Comparison of mapping performance between the CDL product (a) and our mapped results (b) for US in 2020. (c) is enlarged views for typical regions. The left figure of (c) is the aggregated winter wheat map in 1km from CDL data (USDA-NASS, 2023), and the right is our result.

The proposed fraction mapping model can achieve similar spatial dynamics to the distribution of the GlobalWheatYield4km product. Still, improved performance (Fig. 10). Here, our generated fraction mapping model can elevate the accuracy and efficiency of winter wheat dynamics by distinguishing the differences in cropping intensity at 1 km resolution. Overall, the spatial patterns of our mapped results closely match those of GlobalWheatYield4km in the study area (Fig. 10a), suggesting the fraction mapping framework is capable of exploring winter wheat distribution in global countries. To illustrate the specific information regarding our mapping results, we selected six regions with enlarged views and compared differences between the generated winter wheat maps and GlobalWheatYield4km (Fig. 11b). Because there is a finer resolution in our fraction mapping results regarding winter wheat, which provides more detailed spatial information and fraction differences compared to the referred GlobalWheatYield4km data. For example, sites I-III have easily identifiable land types, such as urban extent and riverway. However, for sites IV-VI, our fraction mapping results have notable distribution differences in space compared to the GlobalWheatYield4km data, especially for low-proportion regions. This deficiency might be highly related to coarse resolution (i.e., 4km) and crop mapping errors. Our results indicate that the proposed crop fraction mapping framework helps generate a robust and effective regression model with good performance.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

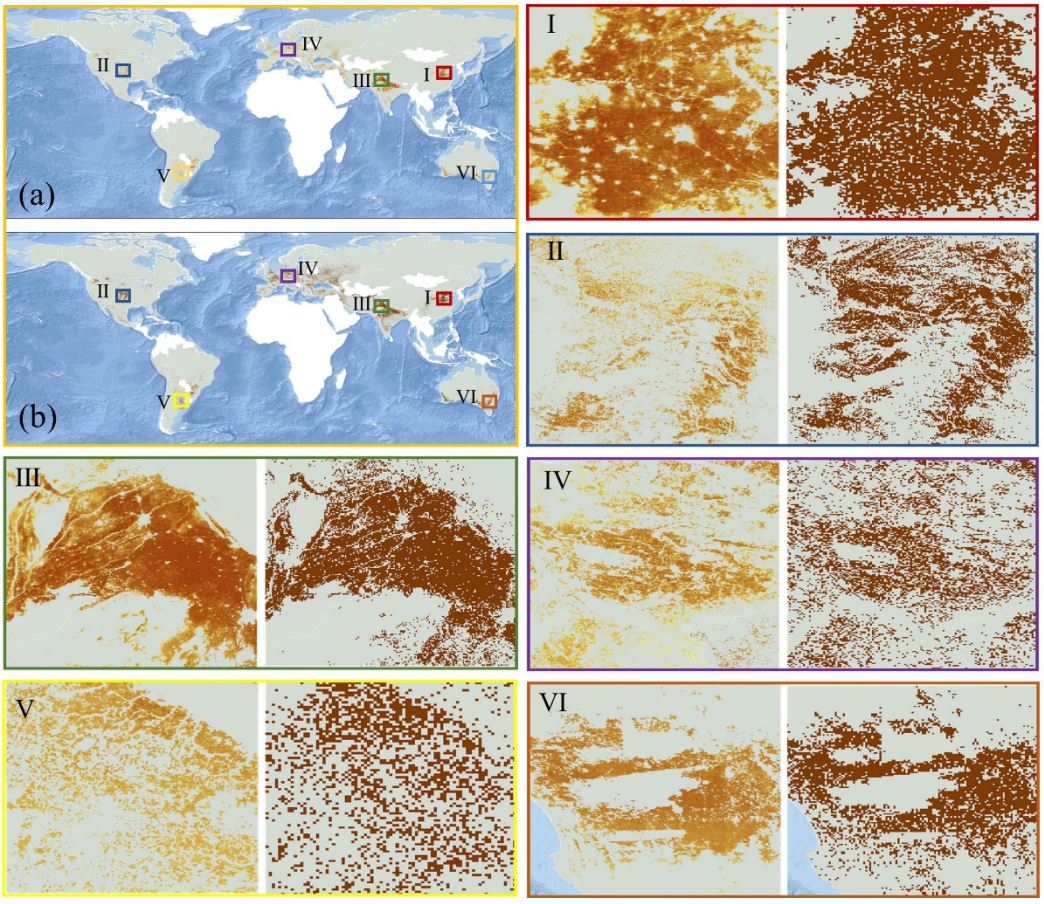


Fig 10. Mapping performance in our study area between our mapped results (a) and the GlobalWheatYield4km product (b) in 2020, including six enlarged views (I-China, II-the US, III- India, IV-Germany, V- Argentina, VI- Australia).

4.3.3. Compared with winter wheat products at various levels

The proposed mapping method is robust in mapping temporal winter wheat distribution dynamics at the country level. Here, we aggregated our mapped result and other winter wheat products (e.g., the 30m winter wheat maps in China, CDL, EUCROPMAP) into 5km (same resolution as the grided yield data) to validate the mapping performance of our method over a long time series. The result indicated that our fractional mapping results had high accuracy in four different regions in 2018, with R-value and RMSE as 0.78 and 0.04, respectively (Fig. 11). Moreover, the proposed fraction mapping framework demonstrated high 1accuracy in producing reliable results across different years (i.e., the mean R-value and RMSE were 0.84 and 0.05 in China from 2016 to 2020, and 0.74 and 0.04 in the US from 2008 to 2020) (Fig. S3, 4). It is worth noting that there are two distinct clusters of data points only in China from 2016 to 2020 (Fig. 11, S3), while other countries exhibit a single concentration of data points at lower values (e.g., the US from 2008 to 2020, and France, Germany, and Poland in 2018) (Fig. 11, S4). That may be attributable to the fact that China has a much higher winter wheat planting density than in

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

the US, due to limited Land resources, smaller cropland scale, and different market demand and economic factors compared to the US (Pan et al., 2021). In the EU, the low winter wheat planting density may be associated with the diversity of agricultural policies and climatic conditions across countries, leading to differences in planting strategies.

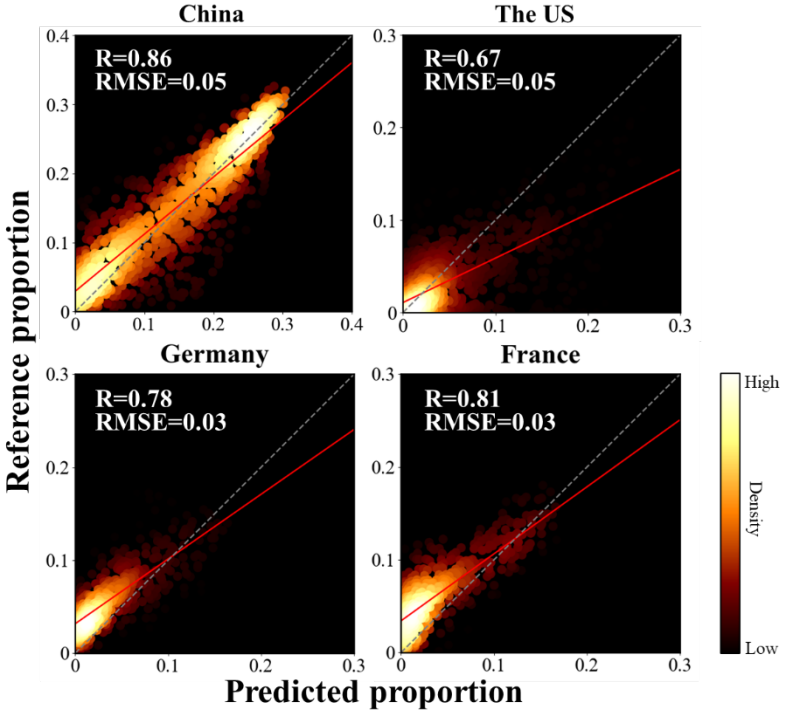


Fig 11. Comparison of aggregated winter wheat fraction data at 5km between our results and the other winter wheat products in 2018.

The employed fraction framework of winter wheat shows great mapping performance regarding the temporal trends of area dynamics at the subregional level, compared to the publicly available winter wheat products (Fig. 12). We selected the subnational-level regions with a relatively high cover of winter wheat in China (i.e., 24 provinces) and the US (i.e., 37 states) to illustrate the effectiveness of mapping performance across time in our study, respectively. The R^2 in both China and the US in temporal dynamics are above 0.95 and 0.72, respectively, suggesting the regression results in those sub-administrative regions are considerably consistent with the reference data (Fig. 12, a-b). For example, the mean R^2 in China and the US is 0.97 and 0.80, respectively. Here, we selected 2016 and 2008 as the start years to determine accuracy in China and the US, considering the conterminous national extent of winter wheat. In addition, the EU countries also have a high R^2 between our results and EUCROPMAP data in 2018, with a high R^2 of 0.88 (Fig. 12c). Given that only a few states in Canada grow winter wheat, we compared the accuracies between our results and ACI data by using yearly national planted area from 2011-2020, which shows our results match well with the survey data (Fig. S5). Overall, our results show a high consistency

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

between predicted results and publicly available winter wheat products, indicating the proposed fraction mapping framework was reliable for mapping winter wheat dynamics.

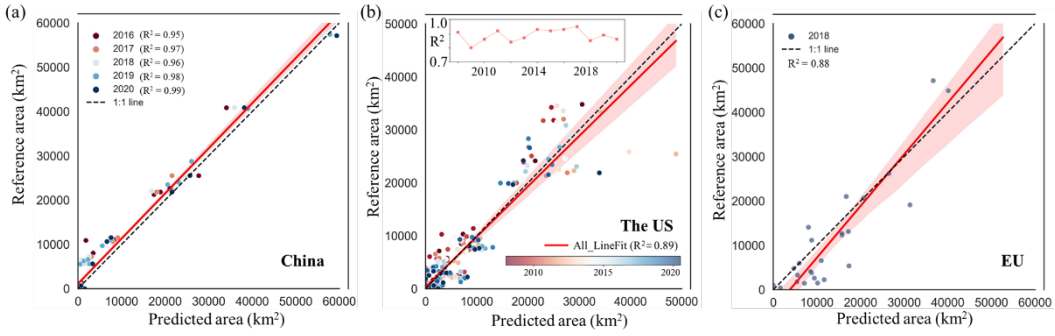


Fig 12. Comparison of region-scale winter wheat areas in China (a), the US (b), and the EU (c). The line chart in (b) shows the changes in R^2 values for the US from 2008 to 2020.

4.4. Uncertainty

Several sources of errors and uncertainties limited the performance of the proposed large-scale fraction mapping approach for winter wheat. First, we adopted two masking methods (i.e., cropland and ROC masks) to exclude non-winter wheat pixels after generating regressed fraction maps. Consequently, pixels with extremely low fractions were excluded in our mapping results, given that our coarse resolution data (e.g., MODIS) are not able to characterize highly spatially mosaic maps derived from fine resolution data (e.g., Sentinel or Landsat). Second, the referred high-resolution products (e.g., 10m to 30m) may also contain errors and uncertainties from crop datasets and spectrum, although we have already excluded many uncertain samples using "adaptive strategies" approach (Wen et al., 2022). For instance, these referred high-resolution productions at a fine spatial resolution are associated with noticeable errors in the fragmented planting region. Nevertheless, the proposed approach can indicate the changes at the global primary planting areas well, which helps support policy decisions and crop growth models.

5. Data availability

The product is publicly available at https://doi.org/10.6084/m9.figshare.32149033 (Wen et al., 2026). The dataset includes GeoTIFF images, with values ranging from 0 to 100 representing the proportion of winter wheat cultivation in each pixel. We encourage users to independently verify the global winter wheat maps.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

6. Conclusion

Here, we developed a framework to map the spatiotemporal dynamics of winter wheat fractions across major global planting regions with robust performance, based on parameters derived from a gaussian mixture model. Specifically, in the main global planted countries, the mean R^2 over the 20-year period is about 0.81 compared with the acreage statistics from FAO data. Moreover, our fractional mapping results in our study show strong agreement with reference datasets in the study area, capturing a similar annual trend from 2001 to 2020. On the one hand, most growing districts (i.e., China, the US, and EU countries) of winter wheat have relatively similar spatial distribution between our mapping results and the crop type datasets. On the other hand, the proposed fraction mapping method appeared significant performance at the regional scale, with mean R^2 in China and the US are 0.97 and 0.80. At the grid scale, the mean R-value and RMSE were 0.84 and 0.05 in China from 2016 to 2020, and 0.74 and 0.04 in the US from 2008 to 2020.

Accurate and timely winter wheat mapping is essential for supporting global food security and promoting sustainable agricultural practices. This is first attempt at mapping global wheat distribution and dynamics over such a long period. The proposed mapping framework is flexible and extendable to other crop types (e.g., maize and rice), showing good performance with existing local maps and statistics (Olofsson et al., 2014). Notably, our fraction map of winter wheat can improve the model simulation ability regarding the global crop growth model (e.g., APSIM). In addition, the generated fine-scale global winter wheat map time series data are highly needed in global crop models, supporting the "Zero Hunger" (Zhang et al., 2022a).

Author contributions

YW, TC and XL designed the research and experiments, performed the experiments, and wrote the original manuscript. TB, KY, LZ, HC, ML, XH, SL, SM, and JH revised the paper. All authors reviewed and revised the manuscript.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (Project No. 2023YFD1501300), the National Natural Science Foundation of China (42101418 & 42371413), and the NSFC Funds for International Cooperation and Exchange (42361164614).

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

References

FAOSTAT: https://www.fao.org/faostat/en/#data, last access: 11 February 2025.

AAFC: https://open.canada.ca/data/en/dataset.

USDA-NASS: https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php.

Ashourloo, D., Nematollahi, H., Huete, A., Aghighi, H., Azadbakht, M., Shahrabi, H. S., and Goodarzdashti, S.: A new phenology-based method for mapping wheat and barley using time-series of Sentinel-2 images, Remote Sensing of Environment, 280, 113206, 2022.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S., and Hostert, P.: Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany, Remote Sensing of Environment, 269, 112831, 2022.

Borrelli, P., Panagos, P., Alewell, C., Ballabio, C., de Oliveira Fagundes, H., Haregeweyn, N., Lugato, E., Maerker, M., Poesen, J., and Vanmaercke, M.: Policy implications of multiple concurrent soil erosion processes in European farmland, Nature Sustainability, 6, 103–112, 2023.

Chen, X., Wang, D., Chen, J., Wang, C., and Shen, M.: The mixed pixel effect in land surface phenology: A simulation study, Remote Sensing of Environment, 211, 338–344, 2018.

Ding, Y., Qu, Y., Peng, Z., Wang, M., and Li, X.: Estimating surface albedo of Arctic sea ice using an ensemble back-propagation neural network: Toward a better consideration of reflectance anisotropy and melt ponds, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–17, 2022.

Dong, J., Fu, Y., Wang, J., Tian, H., Fu, S., Niu, Z., Han, W., Zheng, Y., Huang, J., and Yuan, W.: Early-season mapping of winter wheat in China based on Landsat and Sentinel images, Earth System Science Data, 12, 3081–3095, 2020.

Ghaffari, O., Valadan Zoej, M. J., and Mokhtarzade, M.: Reducing the effect of the endmembers' spectral variability by selecting the optimal spectral bands, Remote Sensing, 9, 884, 2017.

Ghassemi, B., Dujakovic, A., Żółtak, M., Immitzer, M., Atzberger, C., and Vuolo, F.: Designing a European-wide crop type mapping approach based on machine learning algorithms using LUCAS field survey and Sentinel-2 data, Remote sensing, 14, 541, 2022.

Guan, H., Huang, J., Li, L., Li, X., Miao, S., Su, W., Ma, Y., Niu, Q., and Huang, H.: Improved Gaussian mixture model to map the flooded crops of VV and VH polarization data, Remote Sensing of Environment, 295, 113714, 2023.

Hao, P., Di, L., Zhang, C., and Guo, L.: Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples, Science of The Total Environment, 733, 138869, 2020.

Hao, S., Ryu, D., Western, A., Perry, E., Bogena, H., and Franssen, H. J. H.: Performance of a wheat yield prediction model and factors influencing the performance: A review and meta-analysis, Agricultural Systems, 194, 103278, 2021.

Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., and Rautaray, S. S.: A comprehensive survey and analysis of generative models in machine learning, Computer Science Review, 38, 100285, 2020.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Hu, Q., Yin, H., Friedl, M. A., You, L., Li, Z., Tang, H., and Wu, W.: Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates, Remote Sensing of Environment, 258, 112365, 2021.

Huang, X., Huang, J., Li, X., Shen, Q., and Chen, Z.: Early mapping of winter wheat in Henan province of China using time series of Sentinel-2 data, GIScience & Remote Sensing, 59, 1534–1549, https://doi.org/10.1080/15481603.2022.2104999, 2022.

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., and Rowland, C. S.: High resolution wheat yield mapping using Sentinel-2, Remote Sensing of Environment, 233, 111410, 2019.

Jiang, T., Wang, B., Xu, X., Cao, Y., Li Liu, D., He, L., Jin, N., Ma, H., Chen, S., and Zhao, K.: Identifying sources of uncertainty in wheat production projections with consideration of crop climatic suitability under future climate, Agricultural and Forest Meteorology, 319, 108933, 2022.

Jin, S., Zhang, F., Zheng, Y., Zhou, L., Zuo, X., Zhang, Z., Zhao, W., Zhang, W., and Pan, X.: CSKNN: Cost-sensitive K-Nearest Neighbor using hyperspectral imaging for identification of wheat varieties, Computers and Electrical Engineering, 111, 108896, 2023.

Júnior, R. de S. N., Ewert, F., Webber, H., Martre, P., Hertel, T. W., van Ittersum, M. K., and Asseng, S.: Needed global wheat stock and crop management in response to the war in Ukraine, Global Food Security, 35, 100662, 2022.

Kathirvelu, K., Yesudhas, A. V. P., and Ramanathan, S.: Spectral unmixing based random forest classifier for detecting surface water changes in multitemporal pansharpened Landsat image, Expert Systems with Applications, 224, 120072, 2023.

Kempeneers, P. and Soille, P.: Optimizing Sentinel-2 image selection in a Big Data context, Big Earth Data, 1, 145–158, https://doi.org/10.1080/20964471.2017.1407489, 2017.

Kimambo, N. E. and Radeloff, V. C.: Using Landsat and Sentinel-2 spectral time series to detect East African small woodlots, Science of Remote Sensing, 8, 100096, 2023.

Liu, L., Xiao, X., Qin, Y., Wang, J., Xu, X., Hu, Y., and Qiao, Z.: Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine, Remote Sensing of Environment, 239, 111624, 2020.

Ma, L., Yang, S., Gu, Q., Li, J., Yang, X., Wang, J., and Ding, J.: Spatial and temporal mapping of cropland expansion in northwestern China with multisource remotely sensed data, Catena, 183, 104192, 2019.

Nelson, K. S. and Burchfield, E. K.: Landscape complexity and US crop production, Nature Food, 2, 330–338, 2021.

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 Sensing of Environment, 148, 42–57, 2014.

Pan, T., Zhang, C., Kuang, W., Luo, G., Du, G., DeMaeyer, P., and Yin, Z.: A large-scale shift of cropland structure profoundly affects grain production in the cold region of China, Journal of Cleaner Production, 307, 127300, 2021.

Potapov, P., Turubanova, S., Hansen, M. C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., and Cortez, J.: Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century, Nature Food, 3, 19–28, 2022.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Pott, L. P., Amado, T. J. C., Schwalbert, R. A., Corassa, G. M., and Ciampitti, I. A.: Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil, ISPRS Journal of Photogrammetry and Remote Sensing, 176, 196–210, 2021.

Qu, C., Li, P., and Zhang, C.: A spectral index for winter wheat mapping using multi-temporal Landsat NDVI data of key growth stages, ISPRS Journal of Photogrammetry and Remote Sensing, 175, 431–447, 2021.

Ren, J., Shao, Y., Wan, H., Xie, Y., and Campos, A.: A two-step mapping of irrigated corn with multi-temporal MODIS and Landsat analysis ready data, ISPRS Journal of Photogrammetry and Remote Sensing, 176, 69–82, 2021.

Ren, S., Qin, Q., and Ren, H.: Contrasting wheat phenological responses to climate change in global scale, Science of The Total Environment, 665, 620–631, 2019.

Rufin, P., Bey, A., Picoli, M., and Meyfroidt, P.: Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data, International Journal of Applied Earth Observation and Geoinformation, 112, 102937, 2022.

Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V., and Ciampitti, I. A.: Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil, Agricultural and Forest Meteorology, 284, 107886, 2020.

Shao, C., Shuai, Y., Wu, H., Deng, X., Zhang, X., and Xu, A.: Development of a spectral index for the detection of yellow-flowering vegetation, Remote Sensing, 15, 1725, 2023.

Shen, R., Peng, Q., Li, X., Chen, X., and Yuan, W.: CCD-Rice: A long-term paddy rice distribution dataset in China at 30 m resolution, Earth System Science Data Discussions, 2024, 1–33, 2024.

Tran, K. H., Zhang, H. K., McMaine, J. T., Zhang, X., and Luo, D.: 10 m crop type mapping using Sentinel-2 reflectance and 30 m cropland data layer product, International Journal of Applied Earth Observation and Geoinformation, 107, 102692, 2022.

Wang, Q., Shi, W., Atkinson, P. M., and Zhao, Y.: Downscaling MODIS images with area-to-point regression kriging, Remote Sensing of Environment, 166, 191–204, 2015.

Wang, S., Azzari, G., and Lobell, D. B.: Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques, Remote Sensing of Environment, 222, 303–317, 2019.

Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., and Lobell, D. B.: Mapping crop types in southeast India with smartphone crowdsourcing and deep learning, Remote Sensing, 12, 2957, 2020.

Wen, Y., Li, X., Mu, H., Zhong, L., Chen, H., Zeng, Y., Miao, S., Su, W., Gong, P., Li, B., and Huang, J.: Mapping corn dynamics using limited but representative samples with adaptive strategies, ISPRS Journal of Photogrammetry and Remote Sensing, 190, 252–266, https://doi.org/10.1016/j.isprsjprs.2022.06.012, 2022.

Wen, Y., Chen, T., and Xuecao, L.: 20-years winter wheat dynamics in global primary planting areas, https://doi.org/10.6084/m9.figshare.32149033.v1, 2026.

Wójcik-Gront, E.: Variables influencing yield-scaled Global Warming Potential and yield of winter wheat production, Field Crops Research, 227, 19–29, 2018.

https://doi.org/10.5194/essd-2026-347
Preprint. Discussion started: 1 June 2026

Yahiaoui, I., Bradaï, A., Douaoui, A., and Abdennour, M. A.: Performance of random forest and buffer analysis of Sentinel-2 data for modelling soil salinity in the Lower-Cheliff plain (Algeria), International Journal of Remote Sensing, 42, 148–171, https://doi.org/10.1080/01431161.2020.1823515, 2021.

Yang, G., Li, X., Liu, P., Yao, X., Zhu, Y., Cao, W., and Cheng, T.: Automated in-season mapping of winter wheat in China with training data generation and model transfer, ISPRS Journal of Photogrammetry and Remote Sensing, 202, 422–438, 2023.

You, N., Dong, J., Li, J., Huang, J., and Jin, Z.: Rapid early-season maize mapping without crop labels, Remote Sensing of Environment, 290, 113496, 2023.

Zhang, D., Fang, S., She, B., Zhang, H., Jin, N., Xia, H., Yang, Y., and Ding, Y.: Winter wheat mapping based on Sentinel-2 data in heterogeneous planting conditions, Remote Sensing, 11, 2647, 2019.

Zhang, H., Du, H., Zhang, C., and Zhang, L.: An automated early-season method to map winter wheat using time-series Sentinel-2 data: A case study of Shandong, China, Computers and Electronics in Agriculture, 182, 105962, 2021.

Zhang, J., Wang, S., Pradhan, P., Zhao, W., and Fu, B.: Mapping the complexity of the food-energy-water nexus from the lens of Sustainable Development Goals in China, Resources, Conservation and Recycling, 183, 106357, 2022a.

Zhang, Q., Men, X., Hui, C., Ge, F., and Ouyang, F.: Wheat yield losses from pests and pathogens in China, Agriculture, Ecosystems & Environment, 326, 107821, 2022b.

Zhang, Z., Luo, Y., Han, J., Xu, J., and Tao, F.: Estimating global wheat yields at 4 km resolution during 1982–2020 by a spatiotemporal transferable method, Remote Sensing, 16, 2342, 2024.

Zhao, Y., Tao, H., He, P., Yao, X., Cheng, T., Zhu, Y., Cao, W., and Tian, Y.: Annual 30 m winter wheat yield mapping in the Huang-Huai-Hai plain using crop growth model and long-term satellite images, Computers and Electronics in Agriculture, 214, 108335, 2023.

Zhong, L., Hu, L., Zhou, H., and Tao, X.: Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US, Remote Sensing of Environment, 233, 111411, 2019.

Zhu, X., Duan, S.-B., Li, Z.-L., Wu, P., Wu, H., Zhao, W., and Qian, Y.: Reconstruction of land surface temperature under cloudy conditions from Landsat 8 data using annual temperature cycle model, Remote Sensing of Environment, 281, 113261, 2022.