Response to reviewer #4:

Many thanks for your thoughtful and valuable comments and suggestions, which are very helpful in improving our manuscript. We have conducted substantial new experiments and analyses to ensure that the study is more comprehensive and rigorous, and our maps are more reliable. Our responses to the comments point-by-point are included below in blue. The corresponding changes in the revised manuscript are shown in purple.

General comment: Mei et al's work mapped the soybean planting areas across China with a high spatial resolution of 10 meters, spanning from 2017 to 2021, provided important information for sustainable soybean production and management, as well as agricultural system modeling and optimization. In this work, authors summarized five methods of mapping crops by remote sensing. The advantages and uncertainties of each method were compared, and a highly effective for accurately mapping crops over a larger region method named combining unsupervised classification and post-classification methods applied in this paper. They accomplished this by Sentinel-2 remote sensing images from the GEE platform with cropland layer and detailed phenology observations. They validated the results with the census data at both county- and prefecture-level, and with the two existing datasets (CDL and GLAD maize-soybean map).

Overall, I find this work to be valuable. However, I have some concerns regarding the robustness from the sparse number of AMSs in SW Zonal IV and uncertainty in quality of satellite imagery. I hope the authors will consider these points and provide further clarification in their responses and/or revisions. Please find my major comments and minor for clarification below.

Reply: Thank you for your positive and constructive comments, which surely encourage us to further enhance our research quality.

To evaluate the variance in mapping accuracy across different regions, we enhanced each sub-zone's accuracy assessment using statistics and samples (see Reply for Comment 2). Zone IV's mapping results achieved a consistency R^2 of 0.69 with county-level statistics, deemed satisfactory (Figure S5). Validation based on samples indicated an overall soybean accuracy of 87.18% in Zone IV, though it exhibited a relatively lower producer's accuracy of 63.89% than that of other sub-zones (Table S1). Although the verification accuracy there are not as good as those in main producing areas, its accuracy is still acceptable. These findings are highlighted in our results, alongside a comparison the differences in accuracy across regions. To fully and positively respond all your valuable and suggestive comments, we also further listed them point by point in the follows.

Major comments:

Comment 1: The text mentions the need for 10-day time series composite images per month, but in certain areas, the average monthly count of clear observations is insufficient to meet this requirement. Can the existing time series composite methods be optimized to accommodate the inadequacy of observational data?

Reply: Yes, we have optimized the time series composite methods as possible. For the areas with lower clear observations, beside the 10-day time series composite, we also conducted a gap-filling method on the composite time series by replacing the observations by the median of three adjacent observations (i.e., previous, current, and subsequent observations), to ensure the integrity of the time series as much as possible. We supplement in the "Data Processing" section:

"In areas with notably limited clear observations, a gap-filling method was conducted on the composite time series. This method involves substituting any given observation with the median value from three neighboring observations (i.e., previous, current, and subsequent observations) to maximize the continuity and completeness of time series."

Naturally, although the 10-day composite time series were generated as far as possible, this inevitably introduces uncertainty at times (such as 2017) and regions (such as the southwest) where there are particularly few clear observations. We added discussion to the "4.2 The uncertainty from image quality" section:

"In areas with quite lower clear observations, despite a gap-filling method was conducted to generate complete 10-day composite time series, higher uncertainty is inevitable. The gap-filling time series might contain duplicate values, which cannot accurately reflect the crop growth process in reality. Obviously, the total number of images available in 2017 over the study areas was significantly fewer than those of other years (Fig.10a1-e1) ... This might explain the lower user's accuracy of soybean in Zone IV compared to other sub-zones (Table S1) and low overall accuracy based on sample verification in 2017 (Table 2)."

Comment 2: The observations per month of satellite imagery in SW Zonal IV are less, and the AMSs in this zonal also only have two sites. Whether it is possible to increase the observational data or phenological data from remote sensing to test the robust.

Reply: Your suggestion is very helpful. In order to test the robustness of mapping in different regions, we supplemented the statistical data validation of partitions and the point validation based on existing data sets, further demonstrating the mapping accuracy of Zone IV. Zone IV's mapping results achieved a consistency R^2 of 0.69 with county-level statistics, deemed satisfactory (Figure S5). Validation based on samples indicated an overall soybean accuracy of 0.87 in Zone IV, though it exhibited a relatively lower producer accuracy of 0.64 than that of other sub-zones (Table S1). Overall, the accuracy of each sub-zone is acceptable despite some variations.

- **The variations in accuracy among sub-zones based on statistics validation:**

"The mapping accuracy in Zone I closely matched county-level statistics, showing high consistency (R^2=0.86). Zones II-IV also demonstrated reasonable agreement (R^2=0.50~0.69), despite relatively lower accuracy due to the scarcer planted areas (Fig. S5). No significant trend deviation from statistics was indicated for the mapping area in Zone I, with slight overestimates for Zone II and III, and underestimates for Zone IV (Fig. S5). These accuracy variations are acceptable, given the challenges in accurately identifying soybeans in regions where they are planted less prevalently. Specifically, maize is more dominant than soybeans in Zone II, while Zone III is characterized by diverse crops and complex planting patterns. Underestimation in Zone IV is possibly due to fewer clear observations in the southwest. Nevertheless, the overall accuracy across the zones is acceptable."

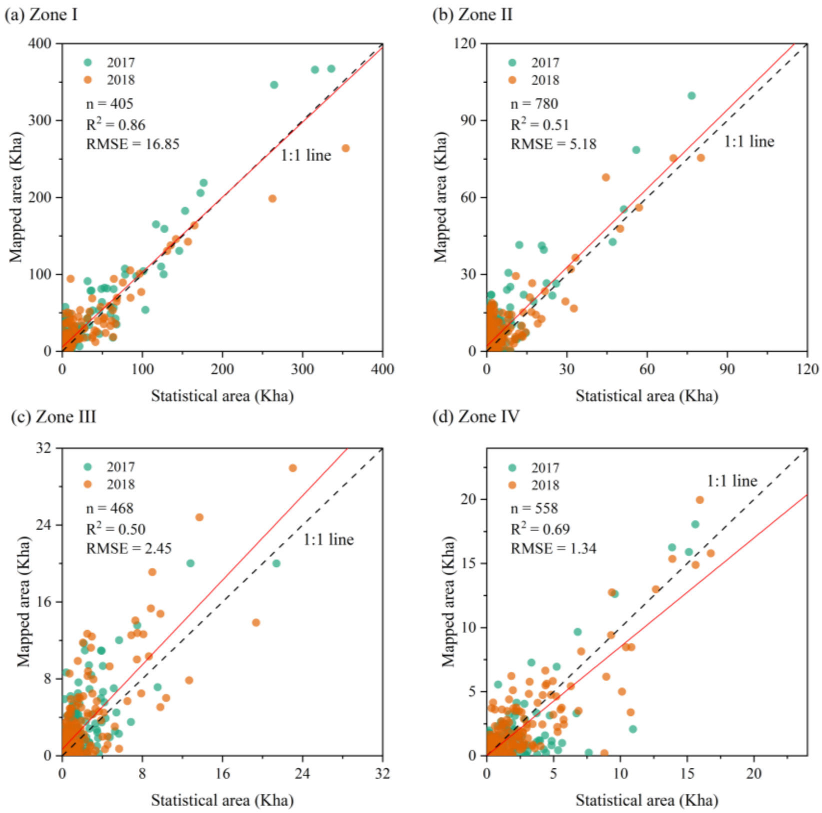

Figure S5. Comparison of soybean areas with county-level statistics in (a) Zone I, (b) Zone II, (c) Zone III, and (d) Zone IV in 2017 and 2018.

➤ **The variations in accuracy among sub-zones based on samples validation:**

"The overall accuracy for each sub-zone in 2019 varied from 83.58% to 90.67% (Table S1). Specifically, Zone I demonstrated the highest producer's accuracy for soybean at 88.31%, aligning with its high consistency with statistics. Zone III achieved the highest overall accuracy at 90.67%, attributed to its superior user's accuracy for soybean, indicating fewer misclassifications, and effective differentiation from non-soybean crops (Table S1). The producer's accuracy in Zone IV was relatively lower at 63.89%, possibly due to the limited samples, high heterogeneity, and fewer clear observations (Table S1)."

Table S1. Confusion matrix of the soybean maps in each sub-zone in 2019.

	Reference	Map		Producer's	User's	F1	Overall
		Soybean	Non-Soybean	Accuracy	Accuracy	Score	Accuracy
I	Soybean	922	122	88.31%	81.09%	0.85	87.12%
	Non-Soybean	215	1358	86.33%	91.76%	0.89	
II	Soybean	233	74	75.90%	86.30%	0.81	83.58%
	Non-Soybean	37	332	89.97%	81.77%	0.86	
III	Soybean	101	26	79.53%	98.06%	0.88	90.67%
	Non-Soybean	2	171	98.84%	86.80%	0.92	
IV	Soybean	23	13	63.89%	92.00%	0.75	87.18%
	Non-Soybean	2	79	97.53%	85.87%	0.91	

Comment 3: To determine the potential cropping areas, authors filtered the pixels exhibiting an EVI maximum value during the growing season greater than 0.4 to remove fallow land. For spatial variation across four zonal, the constant threshold would bring some uncertainty. I expect to see more evidence for selecting 0.4 or a sensitivity analysis of threshold can also be implemented.

Reply: We identified the pixels with maximum EVI values < 0.4 as fallow land because the maximum EVI values for crops are all > 0.4 (except for few outliers) based on all ground samples in 2019 (Figure S1). In addition, studies on crop mapping across China also put forward that EVI values in croplands generally exceed 0.4 at peak growth (Li et al., 2014; Zhang et al., 2017; Han et al., 2022). Thus, using 0.4 as a threshold allows us to strictly remove fallow land. We have provided additional explanations for the threshold choice in the revised manuscript:

"Based on the cropland extracted, we filtered out the pixels exhibiting an Enhanced Vegetation Index (EVI) maximum value during the growing season less than 0.4 to remove fallow land according to the analysis of ground samples (Fig. S1) and previous studies, which found that almost all crops had maximum EVI values above 0.4 (Li et al., 2014; Zhang et al., 2017; Han et al., 2022)."

Reference:

Han, J., Zhang, Z., Luo, Y., Cao, J., Zhang, L., Zhuang, H., Cheng, F., Zhang, J., and Tao, F.: Annual paddy rice planting area and cropping intensity datasets and their dynamics in the Asian monsoon region from 2000 to 2020, Agric. Syst., 200, 103437, https://doi.org/10.1016/j.agsy.2022.103437, 2022.

Li, L., Friedl, M. A., Xin, Q., Gray, J., Pan, Y., and Frolking, S.: Mapping Crop Cycles in China Using MODIS-EVI Time Series, Remote Sens., 6, 2473–2493, https://doi.org/10.3390/rs6032473, 2014.

Zhang, G., Xiao, X., Biradar, C. M., Dong, J., Qin, Y., Menarguez, M. A., Zhou, Y., Zhang, Y., Jin, C., Wang, J., Doughty, R. B., Ding, M., and Moore, B.: Spatiotemporal patterns of paddy rice croplands in China and India from 2000 to 2015, Sci. Total Environ., 579, 82–92, https://doi.org/10.1016/j.scitotenv.2016.10.223, 2017.

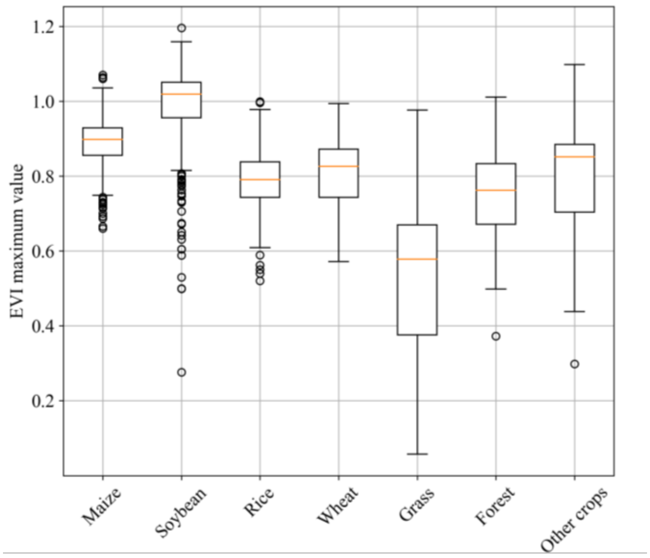

Figure S1. Box plot of the EVI maximum in 2019 based on all ground samples.

Minor comments:

Comment 4: Line 58: "same areas" means the north China?

Reply: Yes, "same areas" here refers to northeast China. We changed it to "three provinces of Northeast China" in the revised manuscript to clarify the meaning.

Comment 5: Line 180, Figure2: The label on the left in Figure2 (i.e. 'Data processing' and 'Accuracy assessment') are set to rotate 180° to match reading habits.

Reply: Thank you for your suggestion. We have angled all the labels on the left side of the diagram for easier reading (Figure 2).

Comment 6: Line 180, Figure2: In step2, part (2) of the dashed box is confusing. What the color represents? If I understand correctly, they represent different layers of indexes. It is recommended to put the abbreviation to the right of the color layers.

Reply: Yes, we have redrawn part (2) of the figure and marked the band or index abbreviations accordingly as you suggested (Figure 2).

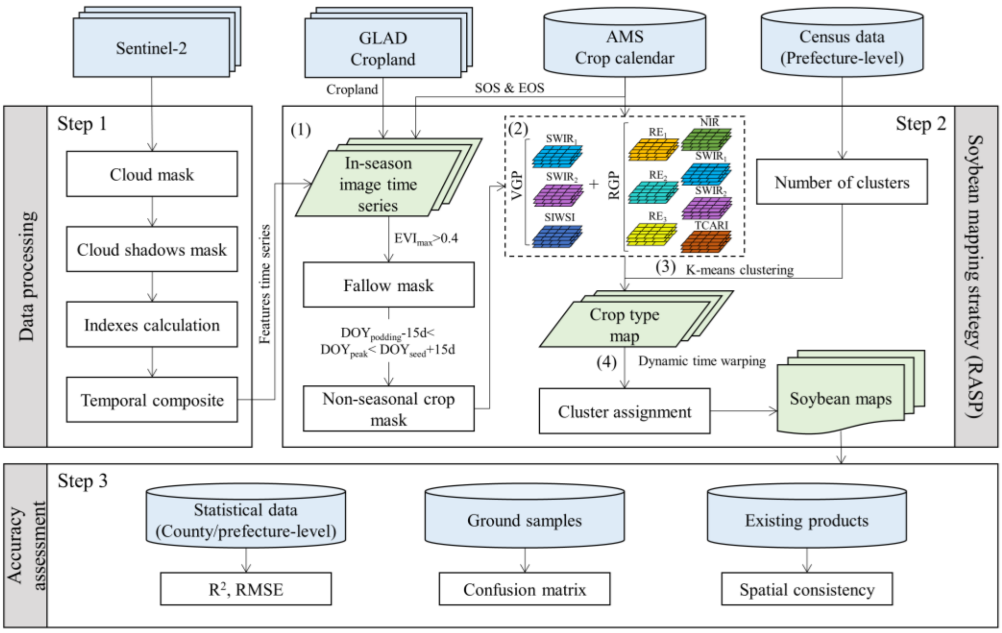

Figure 2. The Regional Adaption Spectra-Phenology Integration methodology for retrieving soybean planting area.