We thank the reviewers for the comments. Our response to the reviewer's comments is marked in blue. The context we cited from the revised clean manuscript with line numbers are marked in green.

Response to referee #2:

Crop type mapping is essential topic as it provides the baseline dataset for crop monitoring community. However such data paper has several issues to be addressed.

We thank the reviewer for the comments. Please see our response below.

1. The necessity of 10-meter resolution within the United States is a subject of debate. Crop classification is a critical component of agricultural monitoring. However, the rationale for developing a 10-meter resolution product over the United States, where large-scale industrial farming is predominant, remains unclear. The incremental benefits of enhancing resolution from 30 meters to 10 meters in such landscapes appear negligible and are not adequately substantiated.

In the context of industrial agriculture in the US, the value of high-resolution (finer than 30 m) crop maps is often ignored, given that the majority of crop fields are large in size. However, as we demonstrated in this study, 10-m maps have many benefits in reducing mixed pixels for many regions. Compared to 30-m maps, these maps can delineate smaller fields clearly (for example, the crop strips in the Appalachian regions in Pennsylvania, and the wetland/field mosaics in North Dakota, as we presented in Figure 13 in the manuscript, also the same figure Fig R.1 below). More accurate field boundaries can be derived from these maps to contribute to field boundary products, thus helping investigations on field-level agricultural practices such as crop rotation.

Fig R.1 The 2022 aggregated 30-m maize and soybean map and CDL show mixed pixels in various landscapes. (a) wetland/agriculture mosaics in North Dakota; (b) center-pivot irrigated fields in Nebraska; (c) strip fields in Pennsylvania. The coordinates of the center points are shown on Imagery © 2025 Google Earth.

We directly demonstrated the benefits of increasing resolution from 30 m to 10 m in the US by comparing our 10-m map with the 30-m CDL. We estimated that the median percentage of mixed pixels across all counties for both maize and soybean was 8%, as illustrated in Line 361-373:

To estimate the number of mixed pixels that might be reduced by improving the CDL's spatial resolution from 30 m to 10 m, we overlaid the 2022 30-m CDL on our 30-m fractional cover map, and computed the number and proportion of CDL maize and soybean pixels that were mixed pixels at the county level (Figure 16). The mixed pixel derived from CDL showed similar spatial distribution patterns to results derived from the aggregated 30-m map (see Figure 14 above). The median percentage of mixed pixels across all counties for both maize and soybean was 8% (Figure 16b, d), which is close to the values of 8% (9% for soybean) derived from our aggregated 30-m map (see Figure 14 above). These consistent mixed pixel estimates from our map and the CDL indicate substantial benefits of 10-m crop maps in reducing mixed pixels over existing 30-m products, especially for regions outside of Midwest, as illustrated in our analyses.

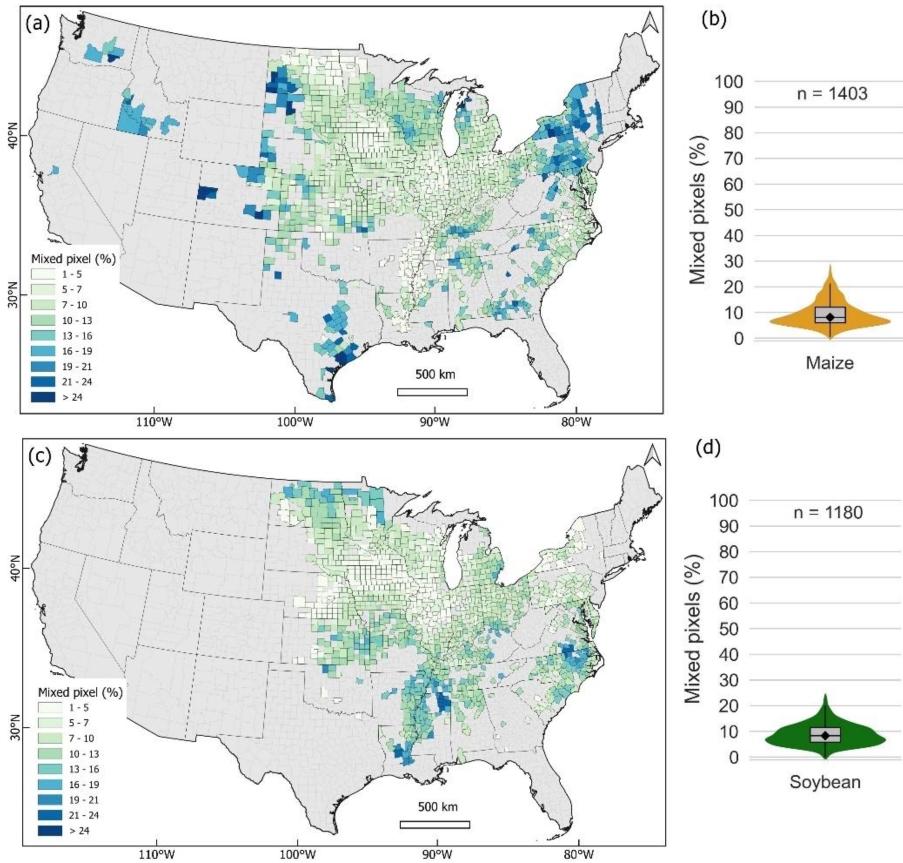

Figure 16: The percentages of 30-m mixed maize and soybean pixels at the county level by comparing our aggregated 30-m map and the 2022 30-m CDL. (a) the spatial distribution of mixed maize pixels; (b) the statistical distribution of mixed maize pixels; (c-d) the same as (a-b) but for soybean. Counties accounting for 99.9% coverage of the national maize and soybean cultivation derived from the 2022 NASS statistics are shown.

2. The absence of competitive advantage in relation to existing products is insufficiently documented. The United States Department of Agriculture (USDA) has already released a 10-meter Cropland Data Layer (CDL) for 2024. In contrast, the present study encompasses only two major crops (maize and soybean) for the 2019–2022 period, whereas the CDL provides nationwide, multi-crop coverage. The manuscript does not demonstrate the unique value or distinct advantage of this two-crop dataset over the publicly available, comprehensive CDL product.

We appreciate the reviewer recognizing the value of CDL. Although CDL also has its own problems, much research relies on this critical dataset that it is difficult to overstate its value. However, we respectfully disagree with the reviewer that because CDL exists, national-scale crop mapping over the United States is a less useful topic to research, for several reasons. First, so far, users of CDL are enjoying the open-access nature of the dataset. However, it is a dataset produced by the United States government, and therefore, it is subject to federal policy regulation. While we are not saying that access to this dataset would be limited in the future, we cannot ignore that future policy change could affect the data access. The most recent example is USDA's Common Land

Units, which was once freely accessible but now restricted. Therefore, it is important to have independent datasets managed by academic institutions. Second, our map products are validated following the best practice guidelines using a stratified random sample and field visit. This is by far the most rigorous procedure for any national-scale crop mapping studies. Our validation method is also different from CDL, but the transparency nature of our validation allows other researchers to gauge the effort level and uncertainty level of implementing stratified random sampling at the country scale.

We have to emphasize that we are not claiming to replace the current CDL with our 10-m crop maps. Our objective is to introduce a robust workflow for crop mapping over large countries, to present a key dataset, and to quantify the benefits of 10-m maps compared to the 30-m products, specifically in mixed pixel reduction. Our methodology does not rely on any existing products for stratification, for training data collection, and for map validation. The comparison with CDL is used only to demonstrate how our 10-m maps improved the crop classification in fragmented fields, on field edges, or along the road networks, even in the context of industrialized agriculture for the US.

Producing a 10-m crop map for all crops in the US is extremely challenging. We are working towards that direction. We believe our established method can be extended to include more crops, although it is subject to more experiments, particularly with regard to rare crops.

3. The added value of this approach is limited in terms of its application for crop area estimation. The estimation of crop area has been demonstrated to be achievable through the implementation of meticulously designed sampling frameworks in conjunction with moderate-resolution classification methodologies (e.g., 30m). The manuscript offers no compelling evidence to substantiate the hypothesis that transitioning from 30-meter to 10-meter resolution would result in a substantial enhancement of area estimation at regional or national scales. This finding calls into question the practical rationale for undertaking high-resolution mapping in this particular context.

Thank you for the comment. We have to emphasize that we focused on generating a high-resolution crop map that could reduce substantial mixed pixels in the US, as we presented in Line 106-107: "The objective of this study is to develop annual 10-m crop maps with Sentinel-2 time series. We also quantify the benefits of 10-m maps compared to existing 30-m products."

We acknowledge that crop estimates are usually derived from well-designed, statistically robust sampling frameworks with the help of satellite imagery at 30-m or 10-m spatial resolutions. We did not state a hypothesis in our manuscript that increasing resolution from 30 m to 10 m would results in a substantial enhancement of area estimation. Quantifying the impact of different spatial resolutions on area estimation is not in the scope of this study; it is an interesting topic requiring more comprehensive research in the future.

4. The temporal scope is limited, and the future plans are ambiguous. The scope of the dataset encompasses the period from 2019 to 2022. In order to enhance the relevance and comparability of the data, it would be advisable to incorporate 2024 data product, thereby facilitating direct benchmarking against the 10-meter CDL dataset for 2024. Additionally, is there any plan to provide annual updates throughout the duration of the Sentinel-2 mission? The annual training and

validation data utilized in the study should be shared publicly, as transparency and reproducibility are paramount in data descriptive paper.

By the time the research was conducted, the sample-based field surveys after 2022 were unavailable. Data pre-processing for Sentinel-2 imagery at the 10-m spatial resolution over the entire CONUS requires tremendous computational resources. Updating such a big dataset would involve a well-planned schedule. We presented our results over multiple years to illustrate the robustness of the method, which could be potentially implemented in operational settings.

When the study was designed and conducted, 10-m CDL were not available. We noticed the significant progress made by the USDA that the 2024 10-m CDL became available in February 2025. We also acknowledge that the 10-m CDL has much improved quality compared to the 30-m CDL 2024, particularly in delineating heterogeneous agricultural fields. If the USDA made 10-m CDL from 2019 to 2022, we could have made a fair comparison with 10-m CDL.

For transparency and reproducibility of our study, we reported that all external data used in this study are openly accessible online, and listed corresponding links in Line 427-435:

"External data used in this study are openly accessible online: 1) the Sentinel-2 data were downloaded from Google Cloud Platform (https://console.cloud.google.com/marketplace/product/esa-public-data/sentinel2); 2) the Cropland Data Layer were downloaded from the US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) (https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php); 3) the TanDEM-X was downloaded from the German Aerospace Center (https://tandemx-science.dlr.de); 4) the agricultural statistics for CONUS were retrieved from the USDA NASS (https://www.nass.usda.gov/Quick_Stats/index.php); 5) the Crop Sequence Boundaries were derived from the USDA NASS (https://www.nass.usda.gov/Research_and_Science/Crop-Sequence-Boundaries/index.php); 6) the road and rail networks were downloaded from the US TIGER database (https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html)."

We understand the concern regarding the field data sharing, but sharing raw field data with precise geolocation info and crop info is not as easy as one might hope. It is especially challenging in the context of the United States. In our many conversations with US farmers, they are often very concerned when data with their crop info are made public while people from geopolitically sensitive regions could use that data. For now, we try our best to make high quality maps and release those map products. These maps are evaluated with the highest standard of accuracy assessment. Other people may make better maps with access to our field data, but the additional utility to data users may be limited.

5. Unclear Methodology for Inter-Annual Consistency. The study uses two types of ground data: "windshield survey" data for classifier training and probability sample data for validation. It is not clearly stated whether new field surveys were conducted every year and whether separate annual models were trained. Without clarification on how phenological variation, seasonal shifts, and survey timing were harmonized across years, the consistency and comparability of the annual maps remain questionable.

Thank you for these comments.

For each year, we conducted the field survey during the peak growing season in July and August. We implemented annual stratified sampling design for each year from 2019 to 2022, as we stated in Line 219-220: "For each year from 2019 to 2022, we selected a separate stratified two-stage cluster sample following the general sampling framework and collected in-season ground data." Our annual maps were validated each year using annual, independent reference data (i.e., we select a new validation sample for each year and collect ground data over the sample in each year). The annual accuracy metrics suggest that the qualities of the annual maps are high and consistent.

More specifically on the methodology, we trained a national model for the multi-year crop classification for all the years. To clarify, we have revised the manuscript to make it clear that all four-year PSU maps are gathered together as training in Line 234-235: "These in-season PSU maps from 2019 to 2022 were then pooled as training labels for national-scale wall-to-wall mapping." We also revised in Line 251-252 to make it clear that one national model across years was trained for maize and soybean classification individually: "Following the approach detailed in Li et al. (2023), we tailored RF binary classifiers separately for maize (RF-Maize) and soybean (RF-Soybean), using the pooled training data from 2019 to 2022."

From the annual field surveys from 2019 to 2022, we collected a large amount of ground data across space and time representing various characteristics across the CONUS. Various factors are reflected from the ground data, including geographical locations, local crop diversities, weather conditions, crop phenological stages, irrigation, etc. These wide-range training data were then used to train generalized models for the national-scale classification.

As we presented in Figure 18 in Section 4.3 (also see the same figure Fig R.2 below), the NDVI for maize and soybean across space and time showed noticeable intra-annual and interannual variations in crop phenology, which were captured by the representative field data collected through our sampling framework. In the same year, 2022, maize and soybean in different states presented various NDVI profiles (a and b). In the same location, the NDVI profiles for maize and soybean also showed variations. Utilizing the multi-year field data for training generalized the model and produced national-scale crop maps with consistent accuracy.

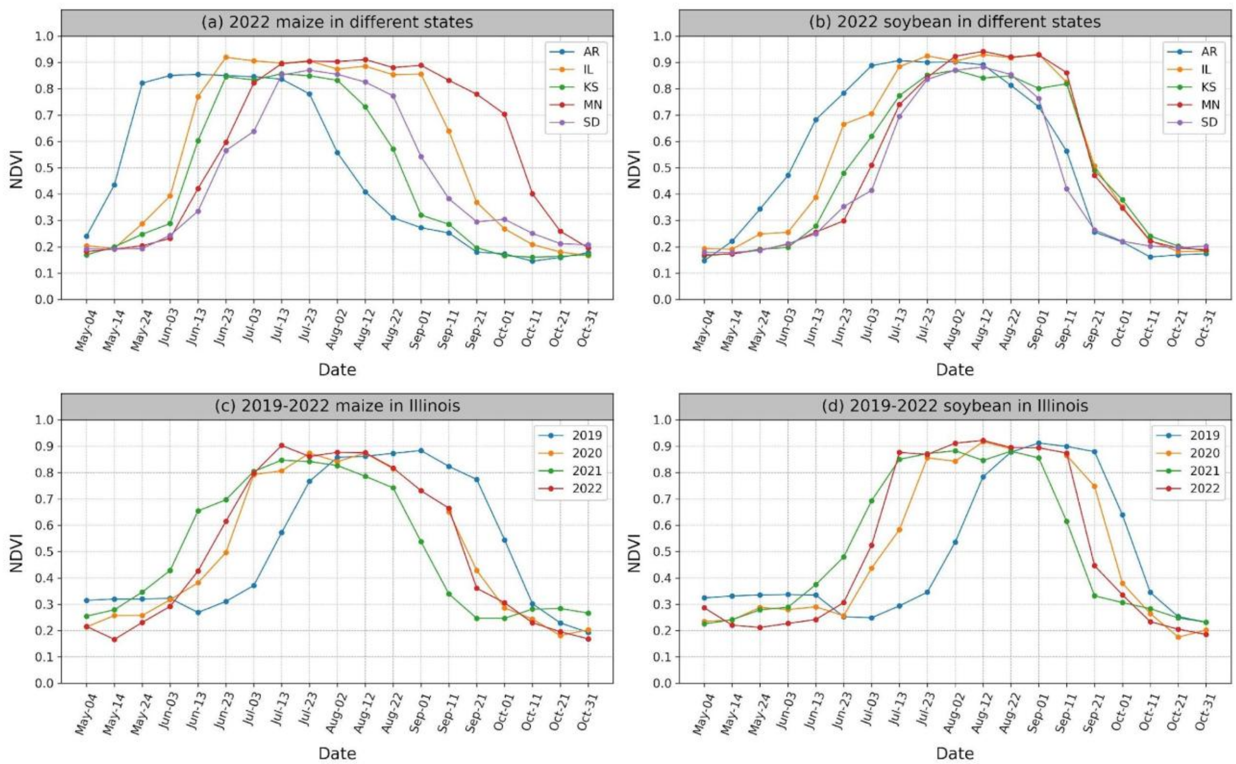

Fig R.2 NDVI time series for maize and soybean from representative sites. (a) 2022 maize NDVI in Arkansas (AR), Illinois (IL), Kansas (KS), Minnesota (MN), South Dakota (SD); (b) the same as (a) but for soybean; (c) 2019-2022 interannual NDVI variations for maize in Illinois; (d) the same as (c) but for soybean. The details about the sites are shown in Figure S4 and Table S6.

To ensure the spatiotemporal consistency in our maps over the entire CONUS, we generated the annual multi-temporal metrics during the growing season from May to October to relative normalize intra-annual and interannual variations in crop phenology, and employed the metrics from 2019 to 22022 as training data to train a generalized model for maize (soybean) classification. As we stated in Line 416-421: "Utilizing the multi-temporal metrics to relatively normalize crop phenological variations, our approach can be applied to generate annual crop maps over large areas, as also illustrated for South America in Song et al. (2021b). Our four-year sampling designs generated large field samples, allowing us to collect representative training data from various growing conditions and geographical regions. Our workflow generated consistently accurate maize and soybean maps over the entire CONUS, from 2019 to 2022. The map accuracies for 2019 (an abnormally wet year), and 2020 and 2021 (both years with normal weather) are consistent with those of 2022."