Response to Review Comments

Ref.: essd-2025-44

Title: CropLayer: A high-accuracy 2-meter resolution cropland mapping dataset for China in 2020 derived from Mapbox and Google satellite imagery using data-driven approaches

Reviewer #1: This manuscript presents a valuable contribution to high-resolution cropland mapping in China through the development of the CropLayer dataset, leveraging data-driven approaches with Mapbox and Google satellite imagery. The integration of deep learning models and active learning strategies to address limitations in existing datasets is methodologically sound. The comprehensive validation against seven existing datasets and official statistics strengthens the credibility of the findings. However, several scientific issues require clarification to enhance the robustness and reproducibility of the work.

Review Comments:

1. The image quality assessment (IQA) using ResNet for cover type classification is innovative, comparative analysis of model performance over other state-of-the-art models for IQA would strengthen this choice.

Response:

Thank you for this valuable suggestion. We adopted ResNet for cover type classification primarily because the task involves only five visually distinctive categories (Planting, Non-Planting, Cloudy, Snow/Ice, and Nodata). While simple thresholding in HSL or Lab color space could distinguish most categories, the Cloudy and Snow/Ice types required additional texture information, for which ResNet provides a robust solution. The model achieved a top-1 accuracy of 95.6%, indicating sufficient performance for this specific IQA task.

We note that more advanced models (e.g., CLIP or Transformer-based frameworks) could potentially offer better semantic understanding and reduce certain residual errors, such as false edges caused by stitching multiple images within a single block. However, such models are typically advantageous in tasks involving hundreds of ambiguous categories and large-scale annotated datasets, which is beyond the scope of this study.

Importantly, our IQA is used to quantify image quality and identify lower-quality blocks for analysis, rather than to “correct” or replace low-quality data. As discussed in Section 5.1, the complementary distribution of low-quality images across Mapbox and Google imagery enables multi-source integration to improve overall mapping accuracy, even when some low-quality images still contribute valuable cropland information.

2. The active learning framework for sample selection mentions "stopping criteria" based on the absence of significant artifacts or underestimation errors, but some quantitative thresholds for termination are not clear, for example, what objective metrics guided the decision to stop sampling?

Response:

Thank you about the comments!

Thank you for this comment. The active learning (AL) framework employed explicit multi-level stopping criteria to determine when sufficient sampling had been achieved. Specifically:

1) Pixel-level: IoU of predicted cropland vs. validation samples must exceed 85%.

2) Block-level: Semantic correctness, evaluated using an XGBoost classifier on validation blocks, must exceed 85%.

3) Provincial-level: Extracted cropland area must reach at least 80% of official provincial statistics.

If any province failed to meet these criteria, additional sampling and retraining were performed iteratively (up to 3 rounds). Provinces where performance could not be improved due to persistent low-quality imagery were excluded. This three-level thresholding ensures that sampling is sufficient for both local accuracy and regional representativeness. We have clarified this in Section 3.2.1 of the revised manuscript.

3. The integration strategy using XGBoost to fuse Mapbox and Google results relies on four feature groups (geographic, IQA, regional attributes, consistency). However, the relative importance of each feature group in improving integration accuracy is not analyzed. A permutation importance analysis would clarify which features drive the model's decisions.

Response:

Thank you for the insightful comment. We conducted a permutation feature importance analysis for the two XGBoost models used in our workflow:

1) Semantic Correctness: using 16 features including 4 DEM features, 12 other.

2) Results Integration: using 28 features including 4 shared DEM features, plus 24 others (each 12 for Mapbox and Google).

The analysis quantifies the relative contributions of different feature groups, including **geographic**, **regional**, **topographic**, **image quality analysis (IQA)**, and **consistency** features. The results demonstrate that

1) Area and AF (Area Fraction) are the most influential features,

2) Regional properties such as Solidity and EN (Euler Number) are critical for identifying correct semantic predictions,

3) Topographic factors such as Slope and Ruggedness play key roles in the integration process.

4) IQA-related features (e.g., Sharpness, Hue, Lightness) show moderate importance in the Results Integration stage, suggesting that imagery quality significantly influences the final decision.

These findings have been incorporated into the revised manuscript (Sections 3.3, 4.3, and 5.2) and summarized in a new figure (Figure 10).

4. The comparison with seven existing datasets shows that CropLayer outperforms others in provincial area estimation, but the reasons for discrepancies in specific regions are not fully explored. Could topographic complexity or cropland fragmentation explain these biases?

Response:

Thank you for the comment. We further analyzed the discrepancies between CropLayer and the eight existing datasets using two complementary metrics: **Area Fraction (AF)** and **Edge Density (ED)**. By constructing two-dimensional histograms of these differences, we identified regions with the largest deviations, which are predominantly located in southern China. These areas correspond to blocks with median slopes of 10-25°, indicating complex topography and highly fragmented cropland patterns. The analysis confirms that regional biases are largely driven by terrain complexity and small, irregular fields, which are challenging to capture even with high-resolution imagery. This result and discussion has been added in Section 4.4 and visualized in Figure12 and 13 of the revised manuscript

5. The Mask2Former model is selected for cropland segmentation based on its highest IoU (88.73%), but the computational efficiency trade-offs (e.g., training time: 11h56m vs. 5h41m (Segformer)) are not discussed. For large-scale applications, model speed and resource requirements are critical.

Response:

Thank you for raising this point. Although Mask2Former requires longer training time (11h56m vs. 5h41m for Segformer), the inference speed is comparable for nationwide predictions, requiring approximately one week for both models. We prioritized accuracy: Mask2Former achieved 88.73% IoU versus 85.10% for Segformer, a substantial difference at the national scale, especially in smallholder and mountainous regions where Segformer consistently underestimates cropland. Therefore, the slightly longer training time was considered acceptable in favor of higher overall mapping accuracy. This clarification has been added to Section 4.2.2.

6. The limitation regarding "inability to capture temporal dynamics" (reliance on 2020 data) is noted, but no feasible path for multi-temporal extension is proposed. For instance, could seasonal imagery from Mapbox/Google (e.g.,2021-2024) be integrated using the same framework?

Response:

Thank you for the comment. The main limitation is that Mapbox and Google imagery lack acquisition metadata and have uneven update frequencies, making nationwide temporal mapping challenging. Urban areas are updated relatively frequently, whereas rural regions may remain outdated for over five years. Nevertheless, once a reliable baseline (2020) map is established, future updates can be implemented incrementally. Specifically, future work could:

1) Apply IQA to newly available imagery (e.g., 2025);

2) Use the existing segmentation and XGBoost integration pipeline to update cropland maps;

3) Incorporate newly released high-resolution datasets such as JLS-5M and ESA PhiSat-2 to improve coverage and accuracy.

This discussion has been added to Section 5.4.

Response to Review Comments

Ref.: essd-2025-44

CropLayer: A high-accuracy 2-meter resolution cropland mapping dataset for China in 2020 derived from Mapbox and Google satellite imagery using data-driven approaches

Reviewer #2: Overall, the study presents a new 2m resolution cropland dataset (CropLayer) which is a valuable contribution given the fine spatial resolution. However, several major concerns should be addressed regarding the definition of cropland, sampling design, methodological innovation, validation approach, and substantive discussion of advantages brought by the high resolution.

Review Comments:

1. The author mentions significant discrepancies among existing cropland datasets and between datasets and statistical data but overlooks the fact that differences can arise from both varying definitions of "cropland" and classification errors. These two aspects should not be conflated. When developing the 2m CropLayer, the paper should explicitly state which definition of cropland is adopted (e.g., FAO, Ministry of Natural Resources, or the GEOGLAM definition). Furthermore, the comparability between the area calculated from CropLayer and statistical area is questionable if their definitions are inconsistent.

Response:

Thank you for this insightful comment. We have clarified the definition of cropland used in CropLayer to ensure conceptual consistency with the official statistical data. In this study, CropLayer strictly follows the criteria established in China's Third National Land Survey (TNLS). According to the TNLS, cropland includes cultivated land used for growing crops such as paddy fields, irrigated land (including greenhouses used for planting), and dry land, as well as land used for temporary crops including medicinal plants, grass, flowers, and trees. It also encompasses newly developed or reclaimed land, fallow land, and areas dominated by crop cultivation, even when interspersed with occasional fruit trees or other vegetation. However, it explicitly excludes orchards, which are separately classified in the TNLS as Plantation land. This definition ensures full conceptual consistency between CropLayer and national statistical data.

In addition to definitional differences, existing publicly available eight cropland datasets in China exhibit substantial inaccuracies when compared with official statistics. In some provinces, reported cropland areas deviate by less than 50% or exceed 200% of the TNLS values, reflecting severe limitations in coverage, methodology, and data quality. These pronounced discrepancies indicate that the challenges are not solely due to varying definitions of cropland, but also stem from inherent errors and limitations in the input data and extraction methods. By adhering to TNLS standards and leveraging high-resolution imagery, CropLayer addresses these deficiencies, providing a more accurate and nationally consistent representation of cropland.

This clarification has been added to Section 2.2.6 (Cropland definition and statistical area) of the revised manuscript.

2. The reliability of visual interpretation for identifying non-planting coverage (e.g., during off-season in mixed cropland-grassland areas) is concerning. It is necessary to clarify how this

challenge was addressed to ensure accuracy.

Response:

Thank you for the comment. We acknowledge that the lack of acquisition metadata in Mapbox and Google imagery complicates distinguishing planting from non-planting coverage. To address this, a ResNet-based classifier was used to identify coverage types (Planting, Non-Planting, Cloudy, Snow/Ice, Nodata), rather than relying solely on manual interpretation. High-resolution off-season imagery often provides additional texture cues, helping to delineate field boundaries even in fallow conditions. Furthermore, **coverage type** is only one of **14 features** used in our XGBoost-based integration of Mapbox and Google results, mitigating the impact of seasonal uncertainty on cropland identification.

3. The design of the sample selection process is not documented. The spatial distribution of samples appears uneven and potentially unrepresentative, with many cropland samples concentrated around a few major cities. A clear sampling framework (e.g., stratified random sampling) should be described to ensure sample representativeness.

Response:

Thank you for the comments. We adopted an active learning strategy for sample selection, which is widely used in semantic segmentation to reduce the high cost of manual labeling. Unlike stratified random sampling, which ensures representativeness only in ***geographic space***, active learning prioritizes regions of high model uncertainty in ***feature space***, thereby improving decision boundaries with fewer samples.

This approach is widely recognized: Settles (2009) provides a comprehensive overview of AL as a core sampling efficiency technique, and more recent studies (e.g., Safonova et al. 2023) demonstrate that AL can reduce annotation costs by 60-99% while maintaining comparable accuracy. Compared with heuristic “trial-and-error” selection or reinforcement learning-based dynamic sampling, AL is particularly suitable for high annotation-cost, **human-in-the-loop** tasks such as large-scale cropland mapping from high-resolution imagery.

The apparent concentration of samples around large cities (e.g., Beijing, Shanghai, Guangzhou) reflects their complex land-use patterns, where cropland is easily confused with features such as golf courses, parks, stadiums or urban vacant land. These areas generated many early misclassifications, so additional samples were added iteratively.

In addition, Guangdong Province (including Guangzhou) contains relatively more samples because it was the first region we completed before expanding to other provinces. Under active learning, the starting area typically requires more samples, as the model rapidly learns cropland patterns during initial iterations. Many generic field patterns captured in Guangdong did not need to be re-collected in other provinces, which makes the overall distribution appear less balanced, though it remained effective for model training.

We have updated the manuscript in two locations to clarify the use of Active Learning for sample selection:

Introduction: added a subsection Optimizing Sample Selection through Active Learning to introduce the rationale and advantages of the method.

Methods, Section 3.2: updated to include detailed **three-level validation schemes, stopping criteria,** and **automated semantic correctness assessment**.”

References:

[1]. Settles, B. (2009). "Active learning literature survey."

[2]. Safonova, A., G. Ghazaryan, S. Stiller, M. Main-Knorn, C. Nendel and M. Ryo (2023). "Ten deep learning techniques to address small data problems with remote sensing." International Journal of Applied Earth Observation and Geoinformation 125: 103569.

4. It is claimed that "independent sample interpretation was conducted." Please explain the specific procedures implemented to guarantee the independence of these samples (e.g., separation of training and validation sets, interpreter blinding protocols, etc).

Response:

Thank you for the comment. We apologize for the unclear wording. The phrase “independent sample interpretation was conducted” refers to the fact that we first selected 3,891 points using systematic sampling, and then separately interpreted these points for Google and Mapbox imagery. In other words, the same set of sampling points was used to generate two independent interpretations corresponding to the two imagery sources, rather than indicating statistical or procedural independence between training and validation sets.

This is also updated in Section 2.2.5.

5. Avoid duplication: Lines 223–226 contain redundant information that should be streamlined.

Response:

Thank you about the comments. Redundant information in Lines 223–226 has been removed.

6. When introducing the seven existing datasets, it would be reasonable to also document their respective definitions of cropland and discuss the similarities and differences among them. This context is crucial for understanding the discrepancies mentioned.

Response:

Thank you for the comment. We have updated Table 3 and Section 2.2.7 to document the cropland definitions for all eight datasets (CACD, CLCD, ESA, ESRI, FCS30, FROM, GL30, SinoLC), specifying the inclusion/exclusion of perennial woody crops and distinctions among paddy, irrigated, and rainfed croplands. Figure 4 now summarizes similarities and differences among these datasets and CropLayer, providing context for interpreting observed discrepancies in provincial cropland estimates.

7. The use of provincial statistical area ratio (>80%) as a conditional check during the extraction process, and subsequent comparison with statistical data for validation, introduces circularity. Since the output is conditioned on the statistics, the resulting high correlation is expected and does not constitute independent validation. Validation against statistical data is therefore of limited reference value.

Response:

Thank you for the comment. We acknowledge the potential for circularity. However, the validation procedure is multi-modal, including pixel-level IoU, block-level semantic correctness, and provincial area consistency. Meeting all three criteria ensures both structural and regional reliability, reducing the risk that overfitting with provincial area that drive the results. Provincial statistics are used as a stopping criterion rather than as the sole validation metric, and the combination of three levels of validation provides a robust assessment of CropLayer accuracy beyond simple area agreement.

8. The authors appear to apply the existing Mask2Former model directly for cropland extraction without significant modifications or improvements. Please clarify the specific innovation(s) of this study compared to simply applying an existing model.

Response:

Thank you for the comment. The innovation lies in how Mask2Former is applied rather than modifying its architecture. CropLayer is the first nationwide 2m-resolution cropland map for China using high-quality, carefully annotated training data. Key methodological contributions include multi-source integration (Mapbox + Google), active learning for efficient sample selection, and multi-level validation. Together, these steps ensure high mapping accuracy and reliability, which is unattainable by merely applying the pre-trained Mask2Former model to low-resolution or poorly annotated imagery.

9. Given the 2m resolution is sufficient for extracting field boundaries, why did the authors not employ recent parcel-based boundary extraction models to create a vector-based cropland parcel dataset instead of a raster thematic map? A vector dataset at the field level would be far more valuable for applications like crop classification and field-level yield prediction.

Response:

Thank you for the comment. While 2m imagery allows identification of cropland presence, field-level vectorization remains challenging in regions with very small or steep fields (e.g., Chongqing), which would require sub-meter imagery for reliable delineation. The primary goal of CropLayer was to provide a nationally consistent 2m raster map, balancing coverage, data availability, and computational feasibility. CropLayer establishes a foundation for future parcel-level vector mapping, enabling field-level analyses once higher-resolution imagery or instance segmentation techniques are integrated.

10. The meaning of "Areas where neither imagery source was utilized" is unclear. Why were large parts of western China not covered by either imagery source? How were the True Negative (TN) areas generated in these regions?

Response:

Thank you for the comment. “Areas where neither imagery source was utilized” refer to regions where no cropland could be extracted from either Mapbox or Google imagery. These areas, largely in western China, are predominantly non-agricultural (mountains, deserts, sparse vegetation), and were designated as True Negative (TN) regions, reflecting correctly identified noncropland blocks. This explanation has been clarified in Section 4.2.4.

11. The Discussion section is relatively weak. The paper should elaborate more on the advantages and quantitative improvements enabled by the 2m resolution in different regions (e.g., fragmented landscapes, smallholder fields), rather than simply stating that 2m is finer than 10-30m. A more substantive analysis of where and how much the resolution enhances accuracy would strengthen the paper.

Response:

Thank you for the comment. The rewritten Discussion emphasizes the quantitative improvements provided by 2m resolution:

Provincial-level accuracy: CropLayer achieves ±10% agreement with official statistics in 30 of 32 provinces, whereas existing eight 10-30 m datasets achieve this in only 1-9 provinces.

Fragmented landscapes: Block-level metrics (AF and ED) show that coarse-resolution datasets underestimate edge complexity and over- or under-estimate cropland in hilly and mountainous regions (e.g., Yunnan-Guizhou Plateau, Sichuan margins). We also added feature importance for the total of 16 features used.

Slope-specific analysis: Underestimation of edge density is most pronounced in 10-25° slope ranges, highlighting the importance of 2m imagery for terraced and hilly fields.

These analyses are now summarized in Sections 4.3.2, 4.3.3, 5.2, and 5.3 to provide concrete evidence of the advantages of high-resolution mapping.