Author's Thanks:

We sincerely appreciate the reviewer's dedicated time and expertise in critically evaluating our work. The constructive feedback has prompted essential refinements to both the scholarly substance and structural clarity of this manuscript, significantly elevating its academic contribution. Below we provide a systematic point-by-point response to each comment. The italicized content represents the modifications made in the manuscript.

Response to Referee #1

Comment 1:

Eq.2 is the most critical part of this study. If P1=0.1, P2=0.2, P3=0.6, P4=0.7, P5=0.8, would you still classify the pixel as non-paddy because |P1-0.5|=0.4? Assuming that P1 and P2 are in the early seasons (e.g., in May) and P3, P4 and P5 are in the peak or late seasons (e.g., August or September), the classification probability for a paddy pixel is expected to increase as crop progresses, because the satellite signals to differentiate paddy/non-paddy become more distinctive in the late seasons compared to that in early seasons. I think it is more reasonable to determine the pixel as paddy in such scenarios. And how do you decide the category if multiple 't' values are found, for example, P1=0.1 and P2=0.9, then |P1-0.5| = |P2-0.5| = 0.4?

Response 1:

Thank you for this insightful and critical comment. You have raised an excellent point regarding the limitation of the decision rule in Eq.2. We have thoroughly analyzed the possibility you pointed out and have addressed it in the Discussion section of this manuscript.

We acknowledge that while the scenarios you described may occur theoretically, their practical impact on the final mapping accuracy is minimal. Specifically, as the paddy rice cultivated in Northeast China is transplanted rice, it exhibits a distinct identification feature during the early growth stages—namely, the flooding signal. Therefore, situations where the probability of paddy rice presence is low in the early stage (e.g., P1 = 0.1) but high in the peak growth stage (e.g., P5 = 0.8) are extremely rare in practice.

To address such exceptional cases, we applied screening conditions including $|P_i - 0.5| > |P_j - 0.5|$, and $|P_i - P_j| > 0.5, where\ i,j \in [1,\dots,m]$, as well as $|P_i - 0.5| = |P_j - 0.5|$, $where\ i,j \in [1,\dots,m]$, to identify regions where the ARE method may perform suboptimally. The

condition $|P_i - P_j| > 0.5$ was used because such cases may lead to misclassification or omission errors when Landsat images from different phenological stages are used to classify the same pixel using the ARE method. Taking the 2020 mapping results of the study area as an example (Fig. 1), these regions comprise only 198910 pixels, covering an area of approximately 179 km², which represents merely 0.014% of the total study area (1.26×10^6 km²). For these limited areas, the final mapping results for the year were determined by selecting the maximum class probability from the classification results of images acquired across different phenological stages within the same year.

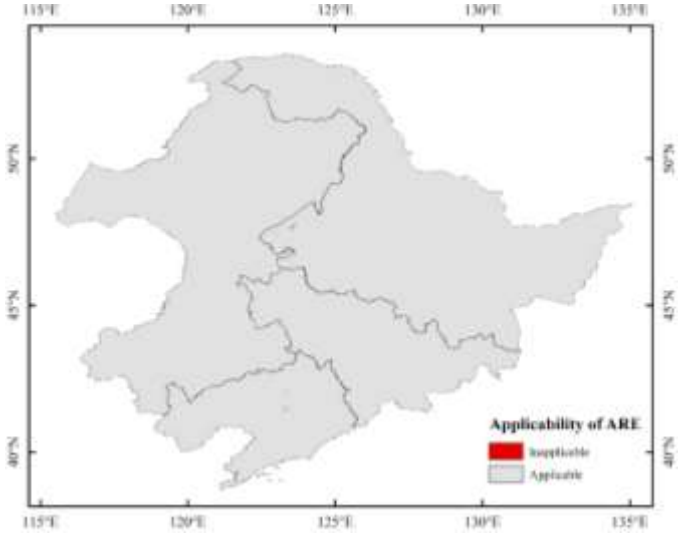

Figure 1 Applicable region of the ARE method

Comment 2:

Line 209: A 30-m x 30-m square is more appropriate than a 30-m radius circle to match the pixels in Landsat. If multiple land cover types co-exist in a mixed sample, how did you determine the final land cover type for validation? Did you use the 50% proportion threshold? This needs to be clarified.

Response 2:

Thank you for your comment. We intended to refer to a 30m × 30m square rather than a 30m radius circle. We acknowledge this error in our original description and have corrected it in the manuscript. Regarding the issue of mixed samples, the final land cover type for validation was determined based on the predominant class within the specified area. Given the large field sizes characteristic of Northeast China, a 50% proportion threshold was deemed highly reasonable for identifying the dominant land cover type. This rationale has been clearly explained in the revised manuscript.

Response to Referee #3

Comment 1:

A key concern raised by other reviewers is about the reliability of the validation procedure. After the authors' clarification, two issues remain. First, the samples interpreted via Google Earth (GE) images might not be entirely accurate, especially since many GE images were acquired in non-growing season. Second, although the random stratified sampling was used to collect the validation samples, the randomness might not be guarantee for each year due to the non-random availability of GE images in each year, which might comprise the comparability of inter-annual accuracy metrics. However, considering the difficulty in obtaining gold-standard validation samples in historical years, I understand these limitations are unavoidable. Nonetheless, I suggest the authors to include some discussions on the remaining issues in the validation procedure.

Response 1:

Thanks for your comment. The samples collected via Google Earth imagery were obtained from images during the growing season (May to September), which helps to mitigate interference from non-growing season conditions. Furthermore, regarding the concern over the non-random availability of Google Earth imagery across years, we have added a discussion of this issue in Section 4.3 of the manuscript and marked the corresponding revisions.

Comment 2:

Inconsistent metric notation: The accuracy metric indicators are inconsistently labeled (e.g., F1 in Table 5 vs. F1 score in Table 6). Please standardize these labels (including UA, PA, OA, MCC, F1) throughout the manuscript.

Response 2:

Thanks for your comment. All relevant indicators throughout the manuscript have been revised for consistency, and the modifications have been marked in this manuscript.

Comment 3:

Line 18: The term 'across-sensors' should be either 'across-sensor' or 'across sensors'.

Response 3:

We thank the reviewer for this comment. The term on line 18 has been corrected to 'across-sensor'. Additionally, we have performed a full-text search to find and correct all similar instances to ensure

consistency. All modifications have been marked in this manuscript.

Comment 4:

Line 398: 'Probability' should be 'probabilities'.

Response 4:

Thank you for the correction. We have changed 'Probability' to 'probabilities' on line 398 (now line 400).