The study aims to generate a 30-m Chinese winter wheat yield dataset (ChinaWheatYield30m) using hierarchical linear model with various input datasets. Results show that the ChinaWheatYield30m dataset was consistent with in-situ measurement data and statistical data, indicated by r (nRMSE) of 0.72** (15.34%) and 0.73** (19.41%), respectively. Overall, there are some aspects needed to be further improved, and the comments are listed below.

1.Line 161-168. How many field-scale yields per year? Where are they distributed? A map is needed for this. And are these fields in the same location each year? If not, a table can be listed to show the number of fields for each year and province.

[Response]: Thank you very much for your suggestion. The sample points may be in the same location each year, or the experimental sampling locations may be increased or modified. In order to more accurately display the sampling points and data in this article, Figure 1 in article has been modified and sample points have been added to the figure. In addition, a table has been added to record the sample numbers of the fields in different years and provinces. We therefore rewrote the following paragraph from Lines 126-128 and table 2:

**Table 2 Detailed statistics on the sample numbers in this study.**

| Province | Anhui | Gansu | Hebei | Henan | Hubei | Jiangsu | Shaanxi | Shandong | Shanxi | Sichuan | Tianji | Xinjiang | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2016 | 12 | 8 | 26 | 45 | | 33 | | 10 | 3 | 11 | 1 | | 149 |
| 2017 | 53 | 4 | 35 | 72 | 16 | 46 | 25 | 59 | 11 | 9 | 1 | 2 | 333 |
| 2018 | 85 | 3 | 63 | 126 | 18 | 47 | 21 | 56 | 14 | 13 | 1 | 3 | 450 |
| 2019 | 85 | 3 | 48 | 130 | 13 | 53 | 17 | 62 | 14 | 10 | | 2 | 437 |
| 2020 | 82 | 10 | 26 | 121 | 11 | 60 | 19 | 52 | 14 | 0 | | | 395 |
| 2021 | 81 | 7 | 25 | 125 | 10 | 26 | 18 | 64 | 8 | 7 | 2 | 3 | 376 |
| Total | 398 | 35 | 223 | 619 | 68 | 265 | 100 | 303 | 64 | 50 | 5 | 10 | 2140 |

2.Line 185-186. Why train the model separately for each province, considering that some provinces have a large geographical span and thus may have internal heterogeneity, I think it may be more reasonable to divide them according to agricultural cultivation subdivisions.

[Response]: Thank you very much for your suggestion. China's agricultural regions refer to the division of China into different areas based on factors such as climate, land use, and crop types, in order to scientifically and reasonably organize and develop agricultural production. The main production areas of winter wheat in China are mainly distributed in some areas of the North China region, the Huang-Huai-Hai Plain, the Yangtze River Middle and Lower Reaches region, and the southwestern region. Each main production area of winter wheat includes multiple provinces. There are significant differences in main crop varieties, crop growth and development, and management practices in these regions. The distribution of China's agricultural regions and provinces can be seen in Figure S1. Figure S1(a) and (b) showed that one agricultural region spanning multiple provinces, for example, the Huang-Huai-Hai Plain includes Henan, Hebei and Shandong. These provinces are a smaller unit of one agricultural region, in other words, each province located in the same agricultural region. Therefore, this article trains the yield model at the province scale to maximize the accuracy of yield prediction results.

Of course, we also agree with your consideration of using agricultural regions for yield prediction. Therefore, we used different agricultural regions as standards for cross-validation of yield results, as shown in Figure 7, Line 236-239 and line 281-293. In addition, we also added a discussion on this aspect.

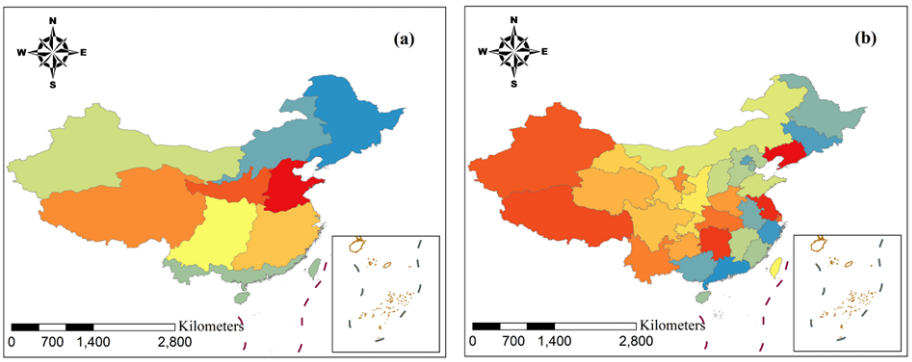


Fig.S1 Map of China's Agricultural Regions (a) and Provinces (b) Distribution.

3.Line 218. Please check, is “relative root mean square error” shorted as “nRMSE” instead of “rRMSE”?

[Response]: Thank you very much for your suggestion. we replaced rRMSE, with nRMSE. The revised sections include all relevant information in the text, figures, and tables throughout the article.

4.Section 3.1 and 3.2, the field-level yield dataset can be further divided into different sets, and then a cross-validation result will better indicate the reliability of the model.

[Response]: Thank you very much for your suggestion. Based on your opinion, in addition to using independent samples for validation, we also selected cross-validation of the model deviation in different agricultural regions. In this paper, commonly used 5-fold cross-validation is used in this study. Modification is incorporated in Line 236-239, Line 281-293 and Figure 7.

5.The province-level validation for this dataset seems to be less meaningful because they are too different in scale and there are too many uncertainties, e.g., crop classification. County or municipal level validation may be more valuable.

[Response]: Thank you very much for your suggestion. The modification has been made according to your opinion, mainly including the acquisition of statistical data, the purpose of comparison, and the comparison results. Modification is incorporated in Line 175-179 and Figure 6.

6.Section 3 is results and discussion, and Section 4 is discussion. I think the title is inappropriate.

[Response]: Thank you very much for your suggestion. We apologized for the mistake. We have changed the title of Section 3 to "Results" and updated and revised the content of the article accordingly.

7.It would be better to present the exact location of the in situ measurement.

[Response]: Thank you very much for your suggestion. We fully understand the purpose of presenting the in situ measurement points, so we have added Table 2 and modified Figure 1 to better display the spatial distribution of the data and sample number.

8.Discussion section is insufficient, especially the uncertainty analysis. For example, how do author deal with these datasets with different spatial resolution, does this bring uncertainties to the final results? Is there any deficiency in HLM model? Etc.

[Response]: Thank you very much for your suggestion. To further discuss the potential uncertainties in the dataset input variables, model, and results, the article analyzed various aspects including dataset resolution, classification dataset, survey samples, and model structure. Modification is incorporated in Line 378-418, as follows:

"*1) Remote sensing and meteorological data used in this study still have uncertainties. This study generated ChinaWheatYield30m dataset with 30-m resolution, the primary reason is we adopted winter wheat classification map from (Yuan et al., ESSD 2020), providing highest resolution of 30-m wheat pixels. The ChinaWheatYield30m input data consist of meteorological variables and remote sensing data, all datasets were resampled to a 30-m resolution to ensure data uniformity. In terms of remote sensing data, resampling Sentinel 2 data to 30 meters may result in loss of some surface information, and the differences between pixels in the image may not be accurately captured. The increase in the number of mixed pixels can lead to uncertainties in yield estimation results. Besides, maximum EVI2 is obtained at the heading or flowering period (Luo et al., 2020), but due to the irregular availability of usable Sentinel 2 and Landsat 8 observations, the maximum EVI2 nationwide may correspond to different phenological periods.In addition, meteorological data is another important component of the yield dataset. To obtain spatially and temporally continuous meteorological driving data, this study utilizes a dataset generated by ECMWF, its meteorological data was timely updated to meet our spatio-temporal demand. However, meteorological data such as precipitation, temperature, and radiation exhibit highly nonlinear and chaotic characteristics (Lorenz, 1993), leading to ongoing debates about the reliability of interpolation methods. The coarse resolution of meteorological data, combined with its high spatial homogeneity over larger areas, weakens its ability to effectively capture the relationship between remote sensing data and yield variations as the second-level correction in the HLM model.*

*2)Uncertainties in winter wheat classifications are transferred to the yield predictions. The wheat classification is based on optical remote sensing data and may be affected by meteorological factors such as clouds and rain (Dong et al., 2020).*

*...*

*4) The uncertainties of HLM application scenarios need further analysis. There is a nested issue between vegetation indices and yield relationships, as well as between meteorological data and yield relationships (Li et al., 2020; Xu et al., 2020). HLM has advantages in addressing this problem. Under similar meteorological conditions, the yield estimation of the model mainly depends on the differences in vegetation indices. In the major wheat production area, variations in crop types, soil types, climate factors, and other factors have an impact on the model's estimation results (Li et al., 2021). The current model only considers the effect of meteorological data on remote sensing yield estimation, and future analyses will incorporate additional factors such as soil to generate more accurate yield datasets. The current model is primarily constructed based on normal production conditions, and estimating winter wheat yield under abnormal climatic conditions introduces significant uncertainties. Therefore, it is necessary to consider stress factors and further improve the framework of remote sensing estimation models for winter wheat in the future.*"

9.Some minor formatting issues: Line 163, "1 $m^2$ per point", Line 197-198, "βmj", "γm0",…, some subscripts are not displayed correctly.

[Response]: Thank you very much for your suggestion. Modification is incorporated in manuscript.