



Random hierarchical model for estimation of wheat yield in the North China Plain at different spatial scales

Xiaobin Xu^a, Wei He^a, Hongyan Zhang^{a,b,*}

^a State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

^b School of Computer Science, China University of Geosciences, Wuhan 430074, PR China

ARTICLE INFO

Keywords:

Yield estimation
Remote sensing
Hierarchical model
Optimal lead time
Parameter importance

ABSTRACT

Accurate and large-scale wheat yield prediction in the North China Plain (NCP) can provide necessary information for agricultural policies and agricultural trade. Many studies have presented wheat yield estimation methods by using existing machine learning methods and remote sensing (RS) or environmental data. However, these methods only blindly input multi-source data into the model and do not consider the hierarchical relationships and interaction between different data types. In addition, there has been scant attention paid to the consistency of yield estimation models across varying spatial scales. To address these problems, a novel dynamic yield estimation model, known as random hierarchical model (RHM), which takes into account the hierarchical relationship of multi-source data, is constructed to estimate the wheat yield in the NCP. First, the time interval of wheat growth period is refined by using the 24 solar terms in China, and a time series multi-source dataset of climate, soil, and RS is constructed. Second, the hierarchical linear model is used to layer multi-source data and randomly select environmental and RS features with multiple time intervals. Multiple hierarchical models are constructed and optimized for integration, and the interrelationships between data collected at different levels are fully utilized, which can improve the accuracy of the yield estimation models in interannual and large-scale applications. Finally, the RHM at different spatial scales is cross-verified by using the measured and statistical data of the NCP for 4 years. The results indicated that the error of the RHM estimation is smaller than that of widely used machine learning models at different spatial scales of field-level measurement data ($R^2 = 0.52$, nRMSE = 16.43%), county-level measurement integration data ($R^2 = 0.62$, nRMSE = 12.83%), and county-level official statistics ($R^2 = 0.68$, nRMSE = 11.41%). Our proposed RHM that considers the hierarchical structure of multi-source data is a reliable and a promising method for improving yield estimation. In addition, heterogeneity in the hierarchical relationships is observed between the different types of data in the RHM at different spatial scales, resulting in differences in the optimal lead time for estimating yield and the importance of key driving factors in the model, indicating that the cross-spatial scale applications of the model are not allowed. This study provides insights for large-scale wheat yield estimation and yield response to different environments and provided evidence and explanation for the prohibition of generalization of models at different spatial scales.

1. Introduction

Understanding the growth and yield of wheat (*Triticum aestivum* L.) is pivotal to managing climate-related risks, assessing national food security, and devising successful food trade strategies (Franch et al., 2015; Lobell et al., 2003). Previous research has been supported the profound influence of climate and soil conditions on crop yields (Alexandrov and Hoogenboom, 2000; Li et al., 2021a; Wang et al., 2016). Enhancing yield estimation accuracy involves integrating these environmental factors into existing systems (Li et al., 2022a; Mathieu and Aires, 2018). The rise

of high spatial and temporal resolution imagery has propelled increased interest in remote sensing (RS), prompting suggestions for crop yield estimation through multispectral and vegetation index (VI) data (Chen et al., 2018; Jin et al., 2020; Kang et al., 2020; Kogan et al., 2013; Wang et al., 2020a). Given the strides in high-performance computing capabilities and diverse algorithms, the utilization of high-resolution, lossless multi-source images encompassing crucial environmental factors such as climate and soil, alongside RS observations, has proven pivotal in facilitating comprehensive wheat yield estimation (Skakun et al., 2018; Wang et al., 2020b).

* Corresponding author at: School of Computer Science, China University of Geosciences, Wuhan 430074, PR China.

E-mail address: zhanghongyan@cug.edu.cn (H. Zhang).

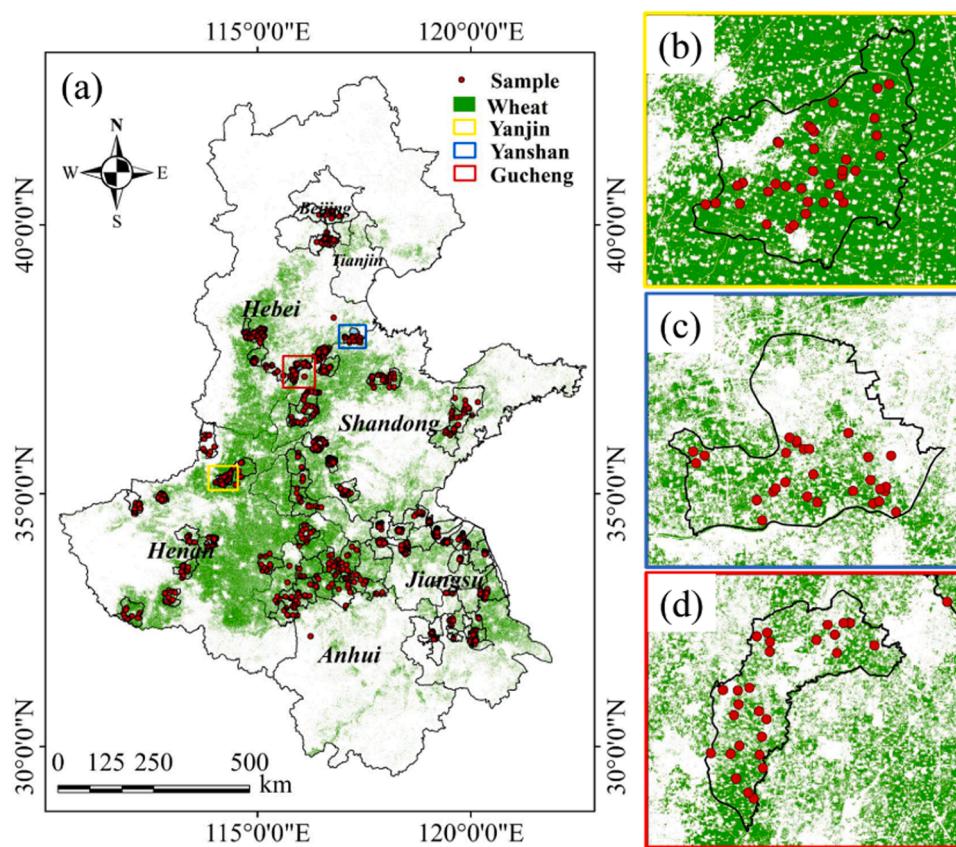


Fig. 1. Geographical location and sample plot distribution map of the study area. (a) Distribution of the wheat planting areas and yield survey points in the NCP. (b) Example of the distribution of the sample plots in Yanjin County. (c) Example of the distribution of the sample plots in Yanshan County. (d) Example of the distribution of sample plots in Gucheng County.

The widely used yield estimation methods currently include three categories: field survey methods, process-based crop physical models, and statistical regression models (Feng et al., 2020b; Li et al., 2021a). While field survey methods persist in official statistical and business yield collection systems, their limitations, including substantial resource requirements and time costs, hinder their sustainable development and timely yield information provision for decision-making needs (Burke and Lobell, 2017; Leroux et al., 2019). The ongoing development of the latter two methods has garnered significant attention to overcome these challenges. The process-based crop physical models provide powerful tools for simulating crop growth, grain formation, and yield estimation (Launay and Guerif, 2005). These models can capture complex interactions between crops and environmental variables and possess strong interpretability (Basso and Liu, 2019). However, crop models have high data requirements, with a large amount of field observation data and complex parameter optimization (Huang et al., 2019; Li et al., 2015). The availability, quality, and cost of data limit its widespread and timely application at the regional scale (Chen et al., 2018; Lecerf et al., 2019). Statistical regression models commonly link yield to chosen crop-related variables or intermediary factors in crop growth processes (Mathieu and Aires, 2018). This approach, characterized by simplicity, cost-effectiveness, and ease of understanding, demands fewer parameter settings (Basso and Liu, 2019). Linear and nonlinear models represent the two primary categories of statistical regression models. The linear regression model provides the most direct mapping relationship between variables, and how to predict the target variable based on the regression formula is relatively easy to understand (Li et al., 2022b; Lobell and Field, 2007). At a small-scale field, establishing a linear relationship between RS variables and target variables can achieve timely prediction (Fu et al., 2022; Zhao et al., 2023). However, ordinary

linear regression models may exhibit significant prediction bias when extended to large-scale and other year applications because the relationship between RS observations and yield varies among different habitats and phenological periods. The addition of multi-source environmental variables does not correct this deviation due to the nonlinear relationship between environment, spectrum, and yield. A possible effective approach is hierarchical linear model (HLM), which can form another level to explain the parameter changes and mixing effects that are not considered in a layer model. Li et al. (2020) and Xu et al. (2020) used this method to combine multi-source data and estimate wheat yield and grain protein content across regions and years, effectively correcting the estimation bias of traditional models. However, they only used three climate variables, namely temperature, light, and rainfall, without considering many other environmental variables that affect the formation of wheat yield. The HLM frequently poorly performs when multiple parameters are selected. Machine learning can be used to handle complex nonlinear relationships with multiple factors and can obtain effective information from multidimensional data to achieve accurate prediction of target variables (Li et al., 2021a; Tian et al., 2021). Previous studies have shown that machine learning methods outperform traditional linear models in estimating large-scale crop yield (Cai et al., 2019; Kaul et al., 2005; Xiao et al., 2019). However, machine learning methods blindly and parallelly input all observation data of different resolutions collected at different levels into a black box, resulting in only one final yield result. Nonetheless, how different variables interact to correct spatial and interannual biases remains obscured. Accordingly, a new yield estimation method, which can balance the advantages of HLM and machine learning, better mine the hierarchical relationships and interactions of environmental variables to improve yield estimation accuracy, and have better comprehensibility in correcting

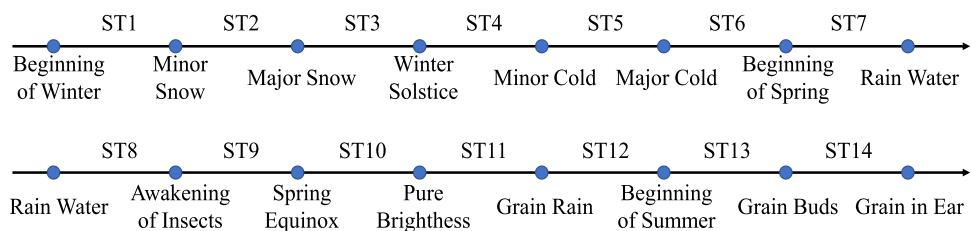


Fig. 2. Time interval divided by 24 solar terms in China. The sequence from ST1 to ST14 represents 14 time intervals during the wheat growing season.

Table 1
Variables used in wheat yield estimation in the NCP.

Category	Variables	Definition	Spatial resolution	Latency
Remote sensing data	EVI	Enhanced vegetation index	500 m	Daily
	NDWI	Normalized difference water index	500 m	Daily
	GCI	Green chlorophyll index	500 m	Daily
Climate data	RH (%)	Relative humidity	0.1° × 0.1°	3 h
	AT (°C)	Accumulated of mean temperature	0.1° × 0.1°	Daily
	WS (m s ⁻¹)	Wind speed	0.1° × 0.1°	Daily
	TE (m of water equivalent)	Total evaporation	0.1° × 0.1°	Hourly
	PE (m)	Potential evaporation	0.1° × 0.1°	Hourly
Soil data	PRE (mm day ⁻¹)	Precipitation	0.1° × 0.1°	Daily
	SR (J m ⁻²)	Solar radiation	0.1° × 0.1°	Daily
	VP(hPa)	Vapour pressure	0.1° × 0.1°	Daily
	STL (K)	Soil temperature level	0.1° × 0.1°	Daily
	SWVL (m ³ m ⁻³)	Volumetric soil water layer	0.1° × 0.1°	Daily

spatiotemporal deviations must be developed based on statistical regression.

The North China Plain (NCP) serves as the primary winter wheat production hub in China, renowned for its dense distribution and high yield (Leghari et al., 2019). A number of wheat yield estimation models based on multi-source data have been constructed in the NCP (Chen et al., 2018; Zhang et al., 2023). However, these methods are determined on a single spatial scale (field scale or county-level scale), and research on the universality of models at different spatial scales is scarce. Agricultural customer groups (including farmers, farm managers, policy makers, and agricultural insurance personnel) have different demands for models at different spatial scales due to data limitations, quality, and confidentiality in China (Azzari et al., 2019). For example, only county-level data are available when estimating the yield at the field scale within a region. However, the spatial heterogeneity of the relationship between the prediction variables and yield in the county may cause serious interference to the performance of established models (Cao et al., 2021). Therefore, the applicability and uncertainty of models at different spatial scales deserve great attention.

In view of this situation, this work proposes a new wheat yield estimation method based on multi-source environmental variables and RS data at different spatial scales (from field scale to county-level scale). First, the climate, soil, and RS data of the entire growth stage of wheat were refined into multiple time intervals without relying on a large amount of uncertainty and difficult to obtain prior phenological information. The features of randomly selecting and combining the optimal time interval alleviate large-scale yield estimation errors. Second, multiple HLMs were constructed by randomly selecting the optimal features in a hierarchical manner, fully taking into consideration the hierarchical

and interactive relationships between multi-source environmental factors that affect the wheat yield formation and RS data that represent yield information. Hierarchical models were optimally integrated and a random hierarchical model (RHM) was constructed to avoid the overfitting phenomenon of traditional hierarchical models. Finally, the RHM yield estimation model was verified by using the 4 year field measured and statistical data in the NCP, and the uncertainty differences in the optimal yield estimation lead time and parameter importance at different spatial scales were evaluated. When the advantages of HLM and machine learning are integrated, this method has good comprehensibility for alleviating large-scale and interannual wheat yield estimation biases. Four years of field-level measurement data (FMD), county-level measurement integration data (CMID), and county-level official statistics (COS) in the NCP are taken as case studies. The main objective of this study is to address two questions: (1) How can we more effectively utilize the hierarchical relationships and interactivity of multi-source data observed at different levels to construct yield estimation models for wheat over a wide range of years without blindly parameterizing multi-source data? (2) How to demonstrate the heterogeneity and uncertainty differences of models established at different spatial scales?

2. Study areas and data

2.1. Study area

The NCP is the second largest plain in China, with a vast area ranging from latitude 32°08'N to 40°23'N and longitude 112°51'E to 122°42'E (Fig. 1). The average altitude is less than 100 m, with flat terrain and numerous rivers and lakes. Various natural soils are fertile, rich in mineral nutrients, and have matured into agricultural soils. Most areas of the NCP have temperate monsoon climate and subtropical monsoon climate with dry and cold winter and hot and rainy summer. Encompassing five provinces—Shandong, Hebei, Henan, Jiangsu, and Anhui—and two megacities, Beijing and Tianjin, the NCP is home to a large population and benefits from convenient transportation networks.

The above-mentioned superior natural conditions and cultural environment make the NCP a major grain production base in China, among which winter wheat is the main winter crop, providing 60% of the wheat output in the country (Duan et al., 2019). The winter wheat in this region is sown in early October of the previous year and harvested in early or mid June of the next year (Wang et al., 2023). However, the smallholder farming system is the main body and major contributor of wheat in the NCP (Cao et al., 2019). Local farmers predominantly employ traditional farming methods for activities such as wheat sowing, field management, and harvesting. Throughout the farming process, they routinely reference the 24 solar terms in China—a cultural legacy born out of agricultural experience and civilization, embodying natural rhythmic changes and phenological laws. For example, the Beginning of Winter (early November) indicates the beginning of winter. At this time, all the wheat in the NCP has been sown and begins to enter the tillering stage. The Grain in Ear (early June) marks a significant increase in temperature and precipitation. At this time, the winter crops in the NCP are harvested, and the summer crops are sown. Chinese farmers have summarized this valuable rule by observing the movement

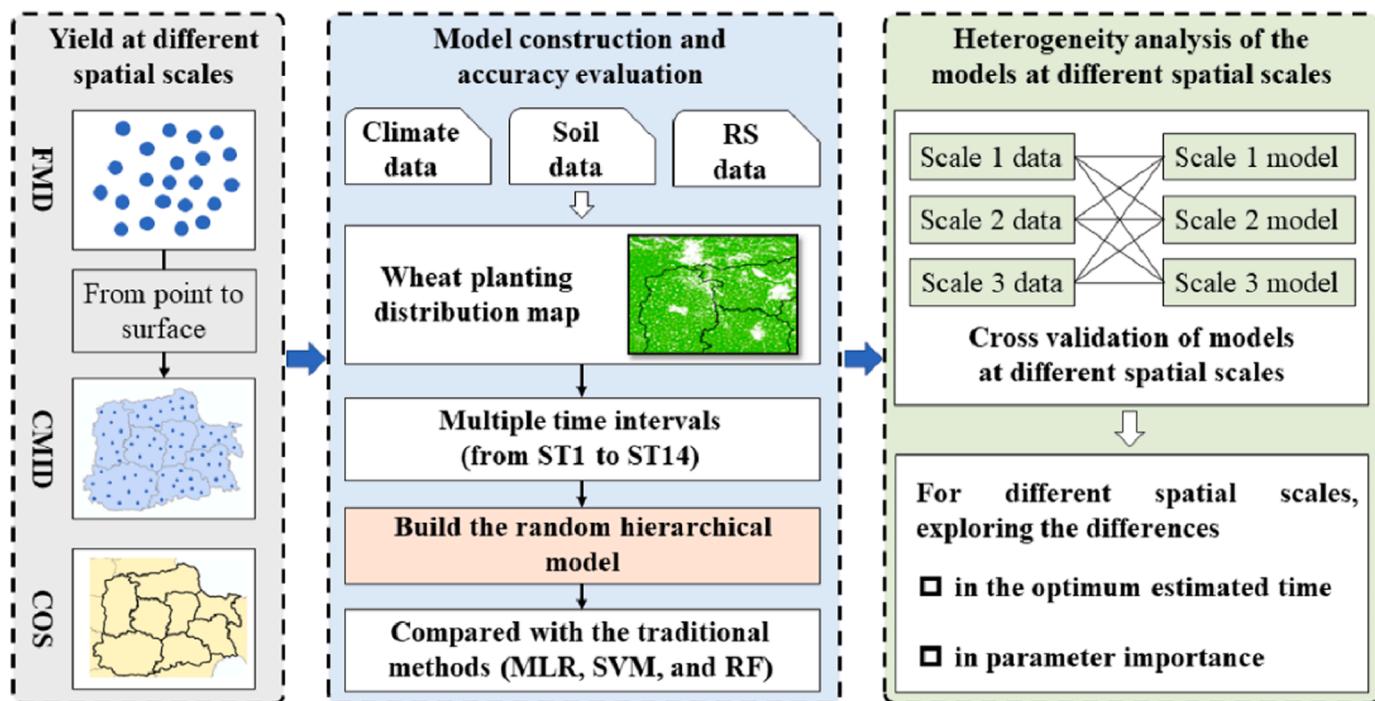


Fig. 3. Overall workflow of this study. FMD: field-level measurement data. CMID: county-level measurement integration data. COS: county-level official statistics. MLR: multiple linear regression. SVM: support vector machine, RF: random forest.

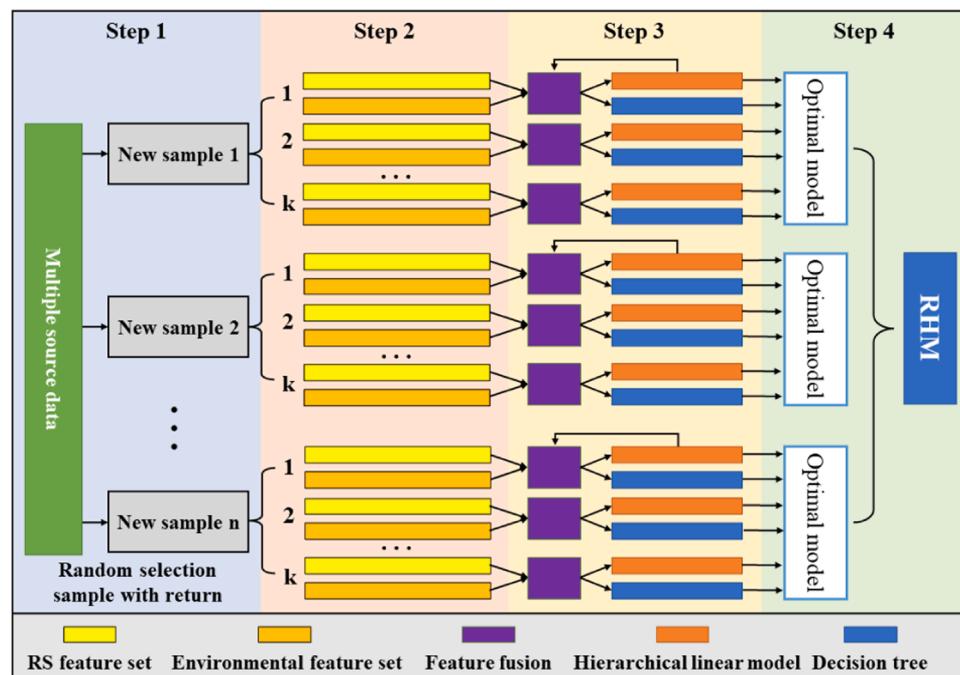


Fig. 4. RHM framework constructed in this study.

of celestial bodies in accordance with the agricultural season and have been officially included in the representative list of the intangible cultural heritage of humanity published by UNESCO (<https://ich.unesco.org/en/RL/the-twenty-four-solar-terms-knowledge-in-china-of-time-and-practices-developed-through-observation-of-the-suns-annual-motion-00647>). Passed down through generations, these criteria traditionally serve as a temporal guide for agricultural production and daily tasks. Currently, providing guidance to farmers are significantly important. In this study, the winter wheat growth season is divided into 14 time

intervals by using 15 solar terms (Fig. 2).

2.2. Data acquisition

Table 1 exhibits a compilation of variables spanning four years (2007–2010 and 2018–2019) in the NCP. These variables were utilized for both the development and validation of wheat yield estimation models, as well as the evaluation of models across various spatial scales. Originating from three distinct data sources, these variables

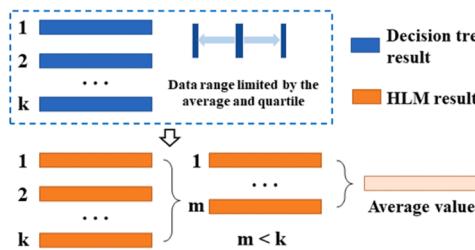


Fig. 5. Schematic of the combination method of the optimal model.

encompassed time-series environmental data (inclusive of climate and soil information), RS satellite imagery, and ground-level data (inclusive of field-measured and county-level official statistical yield data). Detailed information regarding the specifics of the data and their acquisition methods is provided below.

2.2.1. Environmental data

The climate and soil variables for the wheat yield estimates are derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 dataset at the surface level (<https://cds.climate.copernicus.eu>). The temporal resolution of the data is aggregated into daily or hourly scales in the local time zone and corrected to a finer geographic grid (spatial resolution of 0.1°) based on the specific models and integration. The variables provided in this dataset match the input needs of most agricultural and agroecological models (Buontempo et al., 2020; Wu et al., 2021).

Climate promotes and restricts the growth and development of wheat; hence, the climate data can be used for yield estimation (Tian et al., 2021). This study obtained multiple parameters related to water and thermal stress. Relative humidity (RH) describes the amount of water vapor present in the air, expressed as a percentage of the amount required to saturate at the same temperature. The RH was obtained every 3 h a day, and the average RH for each time interval was calculated. The accumulated of mean temperature (AT) was the sum of daily average active temperatures greater than 0 °C during a time interval. Precipitation (PRE) and solar radiation (SR) were calculated as the cumulative sum of daily averages over time intervals, while other climatic parameters in Table 1 were calculated as interval averages.

Soil is the carrier of crop growth and development and the main source of water, fertilizer, gas, and heat for crops (Zhou et al., 2023). It is crucial for crop yield formation and estimation. The soil temperature level (STL) is closely related to crop root respiration, material synthesis, and soil microbial activity. Volumetric soil water layer (SWVL) is

associated with soil depth, groundwater level, and soil texture. The ECMWF Integrated Forecasting System provides four levels of STL and SWVL data products ranging from 0 m to 2.89 m below the surface (i.e., level 1 is 0–0.07 m, level 2 is 0.07–0.28 m, level 3 is 0.28–1.00 m, and level 4 is 1.00–2.89 m). In this study, the daily average STL was calculated, and the STL was the average value of soil temperature in the upper three layers of the data product. The daily average value of SWVL was calculated for the upper three layers of products by using the following formula (Wu et al., 2021):

$$SWVL = \sum_{L=1}^3 SWVL_L h_L, \quad (1)$$

where L is the number of layers of volumetric soil water data product, $SWVL_L$ is the volumetric soil water of layer L , and h_L is the soil depth of layer L .

2.2.2. RS satellite imagery

Satellite VIs contribute to providing and interpreting vegetation-related information and can to some extent alleviate sensor saturation effects. Therefore, satellite VIs have been extensively utilized in research aimed at estimating wheat yield (Cai et al., 2019; Johnson, 2014). In this study, vegetation indices were included in the variables of the wheat yield estimation model. The daily reflectance data were extracted from the MODIS MCD43A4 product with a spatial resolution of 500 m (<https://lpdaac.usgs.gov/products/mcd43a4v061>), and the maximum values of the three complementary vegetation indices (Ma et al., 2021) were calculated for each wheat pixel at each time stage, including Enhanced VI (EVI) (Huete et al., 2002; Ma et al., 2021), Normalized Difference Water Index (NDWI) (Gao, 1996; Taiwo et al., 2023), and Green Chlorophyll Index (GCI) (Gitelson et al., 2005; Wu et al., 2012). The VI representing the county scale was the average value of all wheat pixels in the region, where the wheat pixels were derived from wheat distribution mapping based on a random forest (RF) model, and the monitoring samples were obtained from previous studies (Dong et al., 2020). The VIs were calculated according to the following formulae:

$$EVI = 2.5 \times \frac{(NIR - R)}{NIR + C1 \times R - C2 \times B + L}, \quad (2)$$

$$NDWI = \frac{(NIR - SWIR)}{(NIR + SWIR)}, \quad (3)$$

$$GCI = \frac{NIR}{G} - 1, \quad (4)$$

Table 2

Performance evaluation of the four models constructed by using the non-time phased multisource data and field-level measured data (FMD) during the wheat growth stages. The RS data only include RS variables, environmental data include climate and soil variables, and full data include all the above-mentioned variables.

Year	Error (kg/ha)	RS data			Environmental data			Full data				
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	HLM	RHM
2007–2008	R^2	0.24	0.26	0.26	0.29	0.30	0.27	0.32	0.31	0.28	0.40	0.49
	MAE	1306.66	1253.60	1286.45	1174.39	985.89	1186.87	1054.72	1075.75	1163.71	977.41	907.90
	RMSE	1625.41	1548.24	1574.85	1410.42	1254.45	1492.90	1295.23	1325.46	1471.31	1222.69	1145.51
	nRMSE	25.27%	24.07%	24.49%	21.93%	19.50%	23.21%	20.14%	20.61%	22.88%	19.01%	17.81%
2008–2009	R^2	0.30	0.25	0.30	0.46	0.24	0.28	0.38	0.26	0.27	0.47	0.48
	MAE	1151.43	1387.44	1177.60	1038.17	1460.59	1301.62	1085.25	1398.16	1329.83	975.24	944.24
	RMSE	1467.51	1739.20	1512.16	1313.53	1764.52	1612.24	1394.27	1683.43	1655.45	1269.56	1216.59
	nRMSE	21.06%	24.96%	21.70%	18.85%	25.32%	23.14%	20.01%	24.16%	23.76%	18.22%	17.46%
2009–2010	R^2	0.01	0.02	0.12	0.03	0.02	0.11	0.03	0.33	0.30	0.04	0.34
	MAE	6048.07	2564.37	1992.49	5052.24	2568.21	2018.40	4974.96	1104.42	1144.50	4861.21	1148.62
	RMSE	6580.13	2897.69	2311.04	5623.47	2905.97	2350.67	5515.17	1455.14	1508.50	5424.44	1447.12
	nRMSE	109.17%	48.08%	38.34%	93.30%	48.21%	39.00%	91.51%	24.14%	25.03%	90.00%	24.01%
2018–2019	R^2	0.45	0.41	0.44	0.50	0.44	0.44	0.47	0.46	0.43	0.49	0.52
	MAE	1733.48	1954.57	1790.25	1352.89	1842.30	1888.68	1545.92	1639.51	1902.49	1489.25	1288.35
	RMSE	1954.91	2165.29	2037.22	1607.03	2082.59	2084.67	1797.13	1901.65	2105.47	1700.54	1528.48
	nRMSE	22.98%	25.46%	23.95%	18.89%	24.48%	24.51%	21.13%	22.36%	24.75%	19.99%	17.97%

Table 3
Performance evaluation of the four models constructed by using the time-phased multisource data and FMD during the wheat growth stages.

Year	Error (kg/ha)	RS data			Environmental data			Full data			RHM
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	
2007–2008	R ²	0.17	0.29	0.26	0.00	0.29	0.50	0.01	0.49	0.01	0.55
	MAE	1413.96	1137.72	130.91	7960.90	1049.18	831.80	5528.19	833.91	5324.58	761.29
	RMSE	1813.50	1393.33	1575.70	8204.93	1411.24	1053.82	5876.99	1347.84	5741.02	931.09
	nRMSE	28.20%	21.68%	24.57%	127.60%	21.95%	16.39%	91.40%	20.96%	16.55%	14.48%
	R ²	0.29	0.20	0.29	0.01	0.45	0.44	0.01	0.47	0.45	0.01
	MAE	1250.06	1517.77	1170.65	4954.83	1058.75	1098.80	4621.39	990.46	1073.40	5102.02
2008–2009	RMSE	1571.30	1914.31	1548.98	5482.75	1325.89	1359.50	5227.25	1270.87	1332.21	5526.85
	nRMSE	22.54%	27.46%	22.22%	78.66%	19.02%	19.50%	75.00%	18.23%	19.11%	79.29%
	R ²	0.03	0.22	0.23	0.03	0.15	0.26	0.08	0.31	0.29	0.09
	MAE	2285.03	1435.36	1381.37	2268.69	1742.33	1298.01	2188.85	1164.28	1180.99	1988.97
	RMSE	2798.55	1796.19	1710.38	2783.53	2159.68	1618.71	2751.31	1491.51	1524.29	2555.19
	nRMSE	46.43%	29.80%	28.38%	46.18%	35.83%	26.86%	45.65%	24.75%	25.25%	42.40%
2009–2010	R ²	0.37	0.45	0.41	0.00	0.50	0.46	0.01	0.49	0.47	0.03
	MAE	2081.57	1652.05	1977.44	23321.55	1301.20	1709.27	21030.58	1369.69	1637.77	9502.30
	RMSE	2330.22	1931.66	2208.22	23648.45	1608.33	1921.95	21465.03	1651.57	1848.31	9856.08
	nRMSE	27.40%	22.71%	25.96%	278.06%	18.91%	22.60%	252.39%	19.42%	21.75%	115.89%
	RHM	15.24%	15.24%	15.24%	15.24%	15.24%	15.24%	15.24%	15.24%	15.24%	15.24%

where R , G , B , NIR , and $SWIR$ represent the spectral reflectance of red, green, blue, near infrared, and short-wave infrared, respectively; L represents the adjustment coefficient of the non-green components, such as canopy soil and shadow, and is set to $L=1$; and $C1$ and $C2$ represent the coefficients for correcting atmospheric effects and are set to $C1=6$ and $C2=7.5$. The L , $C1$, and $C2$ parameter values were set according to the MODIS EVI algorithm (Jiang et al., 2008).

2.2.3. Ground-level data

The field-level measurement data (FMD) for wheat yield were collected from extensive field surveys and measurements conducted by the wheat quality alliance and the seed department in the NCP. The distribution of the measured sample plots is shown in Fig. 1, and the wheat yield data per unit area (kg/ha) were calculated for each sample plot. The yield determination method adhered to the principles of a five-point sampling approach, involving destructive sampling within each winter wheat plot. To mitigate edge effects, each sampling point was situated at a minimum distance of 2 m from the field edge, with the area of each sample point recorded (approximately 1 m² per point). After recording the geographical coordinates, samples were brought back to the laboratory where they were standardized to approximately 14% moisture content before being weighed. Subsequently, the yield per unit area (kg/ha) for each sample plot was determined by calculating the average of values derived from five sampling points. The counties with uniformly distributed sample plots (e.g. Fig. 1b–d) were selected, and the county-level measurement integration data (CMID) were obtained by calculating the average value within the domain. The county-level official statistics (COS) for wheat yield were collected from the China Rural Statistical Yearbook (<http://www.stats.gov.cn>), and only the same county and surrounding counties of the CMID were selected to ensure similar geographical environments for comparison. The statistical details of ground-based data can be found in Table A1 in the appendix.

3. Methodology

Fig. 3 illustrates the workflow of this study. First, the yield datasets of the different spatial scales, namely, FMD, CMID, and COS, were established in the NCP by using 4 years of ground field measurement data and statistical data. Second, corresponding to the wheat yield data at different spatial scales, the climate, soil, and RS datasets with multiple time intervals were established based on the 24 solar terms in China. Third, an RHM for estimating wheat yield in the NCP was constructed and compared with the traditional method (i.e., MLR, SVM, and RF). Finally, the heterogeneity of the models at different spatial scales was analyzed through cross-validation to determine the optimal lead time and parameter importance for wheat yield estimation.

3.1. Construction of yield evaluation model

Using single data for large-scale multiyear wheat yield estimation can result in serious estimation bias due to spatiotemporal heterogeneity. Integrating multi-source data, such as climate, soil, and RS, for wheat yield estimation can correct this bias and has been widely proven (Cheng et al., 2022; Feng et al., 2020b; Jiang et al., 2020; Li et al., 2021a). The growth and development process of wheat is influenced by climate and soil conditions, prompting researchers to initially employ meteorological data for large-scale yield estimation. However, meteorological data often exhibit sparse point or coarser resolution raster, with pronounced similarity between adjacent points or pixels. Even within similar growth environments, variations in wheat growth arise from factors like variety and management differences. Consequently, climate and soil data are susceptible to overlooking individual differences within the region, providing only a broad spatial distribution of yield.

The RS information, characterized by higher spatial resolution, serves as the most direct representation of the environmental impact on

Table 4

Performance evaluation of the four models constructed by using the non-time-phased multisource data and CMID during the wheat growth stages.

Year	Error (kg/ha)	RS data			Environmental data			Full data				
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	HLM	RHM
2007–2008	R ²	0.49	0.56	0.50	0.28	0.58	0.50	0.49	0.57	0.58	0.55	0.59
	MAE	1179.64	919.63	1044.66	1654.94	861.44	1080.76	1099.20	865.11	817.04	985.25	809.65
	RMSE	1367.29	1132.39	1302.02	1928.69	1004.39	1275.71	1388.95	1079.98	983.69	1188.23	941.84
	nRMSE	21.47%	17.78%	20.45%	30.29%	15.77%	20.03%	21.81%	16.96%	15.45%	18.66%	14.79%
2008–2009	R ²	0.56	0.48	0.54	0.43	0.59	0.58	0.54	0.53	0.58	0.56	0.60
	MAE	852.06	1066.91	939.66	1129.86	772.54	817.27	928.57	850.86	883.85	850.21	798.31
	RMSE	1024.40	1218.28	1121.89	1347.17	936.31	1002.34	1101.53	1059.36	1011.97	1000.02	927.63
	nRMSE	14.56%	17.32%	15.95%	19.15%	13.31%	14.25%	15.66%	15.06%	14.39%	14.22%	13.19%
2009–2010	R ²	0.31	0.36	0.53	0.09	0.19	0.24	0.10	0.41	0.55	0.13	0.58
	MAE	1131.04	1162.87	797.30	2301.23	1429.10	1610.29	2213.29	1036.63	654.40	2111.02	665.30
	RMSE	1464.50	1389.49	896.15	2708.76	1896.91	1716.73	2607.63	1221.67	819.32	2501.95	775.07
	nRMSE	26.01%	24.68%	15.92%	48.12%	33.70%	30.49%	46.32%	21.70%	14.55%	44.45%	13.77%
2018–2019	R ²	0.53	0.54	0.53	0.48	0.54	0.52	0.48	0.53	0.54	0.53	0.56
	MAE	1931.48	1684.33	1854.52	2113.88	1876.45	1825.62	2134.29	1935.03	1829.39	1911.84	1433.39
	RMSE	2064.55	1900.28	2001.54	2254.17	2038.87	1929.42	2304.01	2083.01	1995.11	2059.23	1700.92
	nRMSE	24.04%	22.13%	23.31%	26.25%	23.74%	22.47%	26.83%	24.26%	23.24%	23.98%	19.81%

wheat growth and development, enabling the differentiation of wheat growth in similar environments. RS information complements climate and soil information but entails observations repeated at another level, wherein one environmental factor may correspond to multiple distinct RS factors. Extensive and multiyear data from various sources create a typical hierarchical nesting relationship, significantly influencing wheat yield estimation in two main aspects. In the first aspect, a set of samples is generated through RS observations of wheat within similar climate and soil grids, establishing a stable intra-layer relationship (as seen in well-performing small-area RS yield estimations). When different climate and soil grids are combined to observe the same sample (e.g., interannual observations), this repeated observation can be considered nested within different contexts of the same sample, giving rise to distinct inter-layer relationships (Li et al., 2020). In another aspect, the relationship between large-scale RS and yield is not notably significant when disregarding the environmental grid. However, within each similar environmental grid, a more pronounced relationship is evident, giving rise to inter-regional inter-layer relationships (Xu et al., 2020). This phenomenon has been underutilized in previous studies for constructing yield models. Given the limitations in aggregation bias, estimation accuracy error, and analysis unit issues (Raudenbush and Bryk, 2001), the traditional methods are unable to analyze hierarchical data and only treat multi-source data as independent and parallel independent variables of the model.

3.1.1. Hierarchical linear model (HLM)

An effective method for parsing hierarchical data is HLM, widely employed in analyzing data with nested relationships, wherein individual observations are nested within higher-level organizational structures (Li et al., 2020). In this study, yield and RS data are individual observations, while the impact of environmental data on RS reflecting wheat yield is considered a higher-level organizational structure. Layer 1 is used to describe the relationship between yield and RS VI changes. The mathematical expression of this layer is expressed using Eq. (5), which includes random effects. Layer 2 is used to study the variation of coefficients and constant terms in the relationship between layer 1 in different environments, as shown in Eq. (6). This two-layer HLM model can provide a more comprehensive understanding of the variation between individual yield and RS VI, as well as how this variation is influenced by environmental factors such as climate and soil.

$$\text{Layer1 : } y_j = a_{0j} + \sum_{n=1}^N (a_{nj} \times V_{nj}) + r_j, \quad (5)$$

$$\text{Layer2 : } a_{mj} = b_{0mj} + \sum_{p=1}^P (b_{pmj} \times C_{pmj}) + \mu_{mj}, \quad (6)$$

where y_j is the dependent variable of the model, j represents the j th sample, a_{0j} is the constant term in layer 1, V_{nj} is the variable in layer 1, a_{nj} is the coefficient of the variable, N is the number of variables in layer 1, and r_j is the random error in layer 1. a_{mj} is a constant and coefficient in layer 1, and is also a dependent variable of layer 2. The value range of m is $0 - N$. b_{0mj} is the constant term in layer 2, C_{pmj} is the variable in layer 2, b_{pmj} is the coefficient of the variable, P is the number of variables in layer 2, and μ_{mj} is the random error in layer 2. The independent variables' parameters in the HLM model should be centered (subtracting the mean) beforehand to diminish numerical computation errors and enhance the numerical stability of the model. The estimation of coefficients in the HLM model can be transformed into an optimization problem. This study employs the method of maximum likelihood estimation to address this issue. Specifically, it aims to maximize the likelihood function, seeking coefficients that maximize the probability of observing the given data.

3.1.2. Random hierarchical model (RHM)

HLM serves as a beneficial tool for handling nested structure data; however, when confronted with multidimensional data at each level containing a substantial number of variables, HLM estimation is prone to overfitting. Additionally, the increase in variables often introduces more noise interference, resulting in poorer estimation accuracy. To address these challenges, this study incorporates the concept of ensemble learning, synthesizing multiple HLMs through the random selection of samples and features, thereby establishing the RHM. The framework of RHM is illustrated in Fig. 4, with specific steps outlined as follows:

(1) Step 1 involves random sampling of input samples. Randomly selecting samples with return from the multi-source sample set, these samples are then reassembled into n sample subsets, each consisting of the same number of samples as the original training set. The random sampling introduces diversity to the samples, aiding the model in developing robustness to variations in the data and reducing the risk of overfitting. The unselected samples were recorded to obtain out-of-bag (oob) data (all unselected samples) while randomly selecting samples multiple times.

(2) Step 2 involves the random sampling of input features. For each sample subset, RS features and environmental features are randomly selected k times, constructing k two-layer feature subsets for each subset. These features are part of all features used to construct sub-models. The selection of RS and environmental features is entirely random and independent, with no interference, and the number of features does not need to be consistent. This ensures that each model is trained on different feature subsets. By randomly selecting features for each sub-model, excessive dependence on specific features is reduced,

Table 5
Performance evaluation of the four models constructed by using the time-phased multisource data and CMID during the wheat growth stages.

Year	Error (kg/ha)	RS data			Environmental data			Full data			RHM
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	
2007–2008	R ²	0.50	0.55	0.52	0.00	0.62	0.58	0.00	0.58	0.60	0.02
	MAE	1105.23	1027.31	1031.59	41179.77	647.97	872.78	7019.19	820.81	732.59	5522.24
	RMSE	1335.87	1165.40	1214.55	41965.51	779.28	1006.40	7912.45	974.33	870.91	6437.48
	nRMSE	20.98%	18.30%	19.07%	658.98%	12.24%	15.80%	124.25%	15.30%	13.68%	101.08%
	R ²	0.59	0.54	0.54	0.00	0.62	0.53	0.00	0.58	0.60	0.01
	MAE	843.17	877.41	917.53	22607.40	797.92	949.24	16122.89	883.02	779.42	8333.24
2008–2009	RMSE	957.94	1083.12	1100.00	26148.79	910.06	1062.08	16683.38	1012.82	924.48	8452.97
	nRMSE	13.62%	15.40%	15.64%	371.77%	12.94%	15.10%	237.20%	14.40%	13.14%	120.22%
	R ²	0.51	0.48	0.36	0.00	0.30	0.10	0.00	0.46	0.53	0.01
	MAE	813.15	869.94	1119.23	1096.78	1214.69	2450.31	8250.16	830.78	709.15	9027.55
	RMSE	935.40	1032.23	1427.92	12368.91	1558.67	2597.37	10740.29	1094.99	865.27	9358.36
	nRMSE	16.62%	18.34%	23.36%	219.71%	27.69%	46.14%	190.78%	19.45%	15.37%	147.5%
2009–2010	R ²	0.39	0.62	0.55	0.00	0.66	0.53	0.00	0.66	0.67	0.67
	MAE	2419.26	1050.99	1822.86	41537.56	878.04	1870.37	11671.21	841.15	826.90	8123.66
	RMSE	2669.41	1298.26	1962.17	42510.27	1120.61	2043.56	12446.56	1082.60	1051.05	8534.82
	nRMSE	31.08%	15.12%	22.85%	494.99%	13.05%	23.80%	144.95%	12.61%	12.24%	99.35%
	R ²	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MAE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2018–2019	RMSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	nRMSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	R ²	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MAE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RMSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	nRMSE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

contributing to enhanced generalization performance. In this study, the number of RS features was set to three or fewer, and the number of environmental features was set to five or fewer after multiple experiments.

(3) Step 3 involves the fusion of two-layer features and the construction of sub-models. Based on the randomly selected two-layer features for each sample subset, RS features and environmental features are fused according to the data structure of HLM (Eqs. (5) and (6)), resulting in the construction of HLM. This process ensures that each sample subset utilizes k different and independent HLMs constructed from diverse features. To facilitate the optimal integration of multiple HLM results in Step 4 to obtain the final yield estimation, a decision tree is concurrently constructed while building the HLM.

(4) Step 4 constitutes an optimal model selection mechanism. In this step, the results of all k sub-models for each sample subset are first individually integrated. Subsequently, the average of the results from all n sample subsets is computed as the final yield estimation value. The integration mechanism for the optimal model within each sample subset is illustrated in Fig. 5. It employs the quartile range estimated from k decision trees to constrain the estimates of multiple HLMs. Estimates falling outside the quartile range are considered unreliable results. These unreliable HLMs are either eliminated or mapped to obtain m HLM estimates ($m < k$). The average of these m HLM estimates is computed as the optimal estimation for the subset. Here, n and k represent two hyperparameters of the RHM. Larger values for n and k indicate a more reliable model accuracy, both set to 5000 in this study.

Through the aforementioned four steps, the wheat yield estimation model of the RHM can be developed. In addition, each sub model trained was tested using oob data to obtain the mean absolute error (MAE). The calculation formula is shown in Eq. (7). The order of the q th feature was randomly shuffled, keeping the other features unchanged, and the MAE' was calculated. Then, the importance score of each feature was calculated using Eq. (8). The relative importance of each feature was obtained by normalizing the proportion of all features. This study developed a wheat yield estimation system at multi-lead time based on the RHM. The RHM could be constructed to estimate the wheat yield at any time stage (ST1–ST14) by adding estimation factors over time. For example, when estimating yield at ST2, the estimation factors of ST1 and ST2 were used as input variables for the RHM method. Therefore, the number of estimation factors for the RHM gradually increased as time advanced.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (7)$$

$$IS_q = \frac{\sum_{p=1}^N (MAE_p - MAE'_{pq})}{N}, \quad (8)$$

where n denotes the number of samples; y_i and \hat{y}_i are the actual and estimated values for the i th sample; \bar{y}_i denotes the average of the actual values; IS is the importance score, q represents the q th feature, N represents the number of sub models, and p represents the p th sub-model.

3.2. Model comparison

Three widely used models were compared to better evaluate the performance of the RHM in estimating the wheat yield in the NCP. These models were MLR, SVM, and RF, which were described in detail below.

MLR is a commonly used linear regression method in the estimation and prediction of production and empirical model or approximation function with more than one regression variable (Mehdaoui and Anane, 2020). Moreover, MLR is the preferred approach due to the simplicity and ease of deployment of establishing empirical relationships (Saad El Imanni et al., 2022).

SVM is a class of generalized linear classifiers that perform binary classification of data using supervised learning (Liakos et al., 2018). It

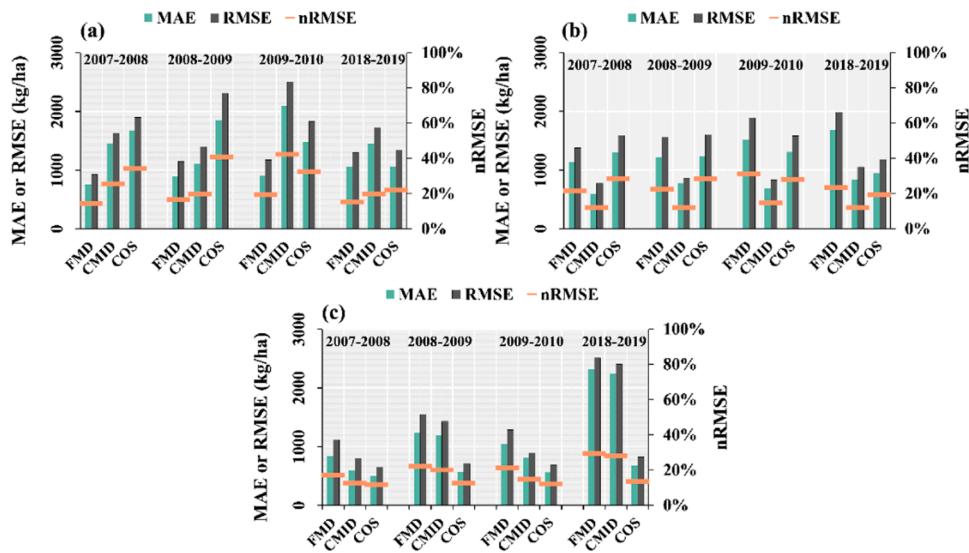


Fig. 6. Comparison of the cross-validation accuracy of the models with different spatial scales. Panel (a) indicates that all three spatial scale models are validated by using the FMD. Panel (b) indicates that all three spatial scale models are validated by using the CMID. Panel (c) indicates that all three spatial scale models are validated using the COS. The horizontal coordinates represent the validation data of the model in the corresponding year.

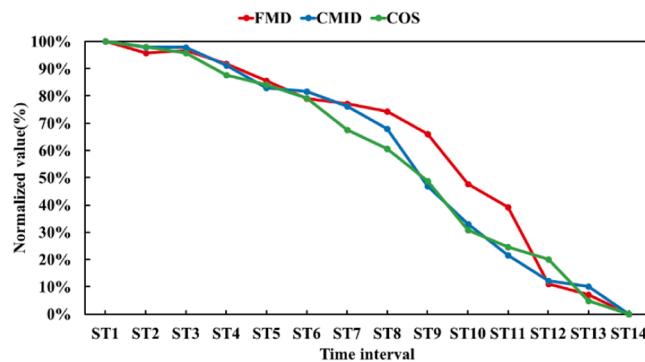


Fig. 7. Normalized values of the performance indicators (nRMSE) for the different spatial scale wheat yield estimation models at different time stages.

establishes a decision boundary through solving a convex quadratic programming problem for the maximum margin hyperplane (Zhao et al., 2022). SVM provides a clearer and more powerful way to learn complex nonlinear equations compared with logistic regression methods (Wang et al., 2023). This research optimized two SVR model hyperparameters—regularization parameter (C) and kernel function parameter (gamma)—using random optimization (Bergstra and Bengio, 2012).

RF, an ensemble learning algorithm, features multiple decision trees constructed in parallel through random sampling of datasets and features (Breiman, 2001). Trees in the forest, all weak and independent learners, combine to enhance generalization performance, especially for weak learners (Wang et al., 2023). RF excels in processing high-dimensional datasets, providing accurate, nonlinear analysis, and resilience against overfitting (Zhou et al., 2022). This study optimized two RF hyperparameters—number of base learners (mtry) and number of attributes per tree (ntree)—using random optimization (Bergstra and Bengio, 2012).

3.3. Performance evaluation

Three widely used estimation methods, namely, MLR, SVM, and RF, were compared to evaluate the performance of the developed RHM. To test the stability of the model between different years, all approaches were evaluated using data from the validation year, while data from

other years were used for model training. For example, when using 2019 data to evaluate the accuracy of the method, the data samples from 2007 to 2010 were used for model training. The R^2 , MAE, root mean square error (RMSE), and normalized RMSE (nRMSE) were used to compare the accuracy of different approaches in estimating the wheat validation sample sets. The MAE can estimate the residuals and demonstrate the uncertainty of the estimated results (Ma et al., 2021). The calculation formulas are shown in Eq. (7). The RMSE is used to calculate the average square deviation between the estimated and the actual values (Feng et al., 2020b). The nRMSE can provide a dimensionless percentage representation of the deviation from the actual and estimated data, facilitating performance comparisons between different models (Xu et al., 2021). The calculation formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (9)$$

$$nRMSE = \frac{RMSE}{\bar{y}_i}, \quad (10)$$

where n denotes the number of samples; y_i and \hat{y}_i are the actual and estimated wheat yields for the i th sample; and \bar{y}_i denotes the average of the actual wheat yield.

4. Results

4.1. Accuracy evaluation of the RHM for the FMD

Wheat yield estimation models were constructed by using the multi-source FMD that has not been resampled into multiple time intervals during the whole growth season of wheat in the NCP. The results are reported in Table 2. All estimation accuracy rates were relatively ordinary. The models constructed with only RS data, environmental data, and full data were compared, and the difference in estimation results was insignificant. The linear models (MLR or HLM) with few feature quantities have similar performance compared with the machine learning nonlinear models (SVM and RF), and the validation accuracy rates in 2008–2009 and 2018–2019 were even better. Among them, the accuracy of the layered HLM model is significantly higher than that of the single-layer MLR. However, the linear model had low estimation accuracy in 2009–2010, resulting in unreliable results. In comparison

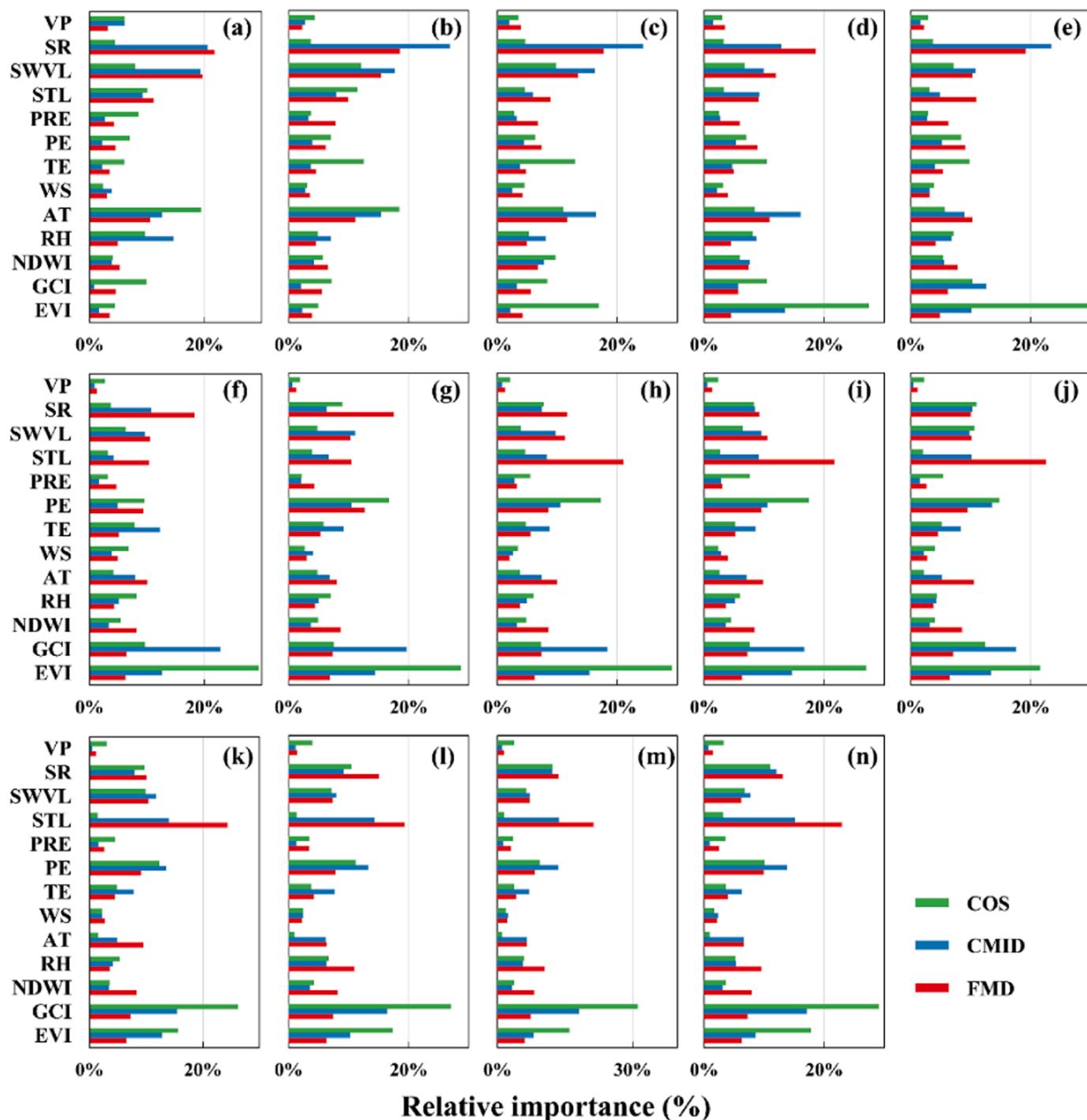


Fig. 8. Relative importance of the different estimation variables at varying time intervals for the three spatial scale models. Panels (a)–(n) demonstrate the importance of the yield estimation model established at 14 time intervals from ST1 to ST14.

with the other widely used models, the RHM developed in this study outperformed all the other models during the 1 year leave for verification, with an average R^2 of 0.46, MAE of 1072.28 kg/ha, RMSE of 1334.43 kg/ha, and nRMSE of 19.31%. The performance of the other models did not significantly differ from each other.

The model estimation results for resampling multi-source data into multiple time intervals are listed in Table 3. All models that only use RS data to estimate yield have no significant change compared with the data modeling without time interval. The accuracy of the models for estimating yield using climate data and full data has significantly improved. The model constructed using full data has the best

performance, followed by the model created using environmental data and the model developed utilizing RS data. The developed RHM is superior to other models, reaching an average R^2 of 0.52, MAE of 908.87 kg/ha, RMSE of 1138.11 kg/ha, and nRMSE of 16.43% over 4 years. The performance of the RF and SVM models was similar. In comparison with the second most accurate model, the estimation error of the RHM was significantly reduced, with an average MAE decrease of 272.65 kg/ha and nRMSE decrease of 4.24%. However, the linear model experienced a serious overfitting phenomenon due to the excessive features generated by the resampling time interval, and the verification accuracy is poor.

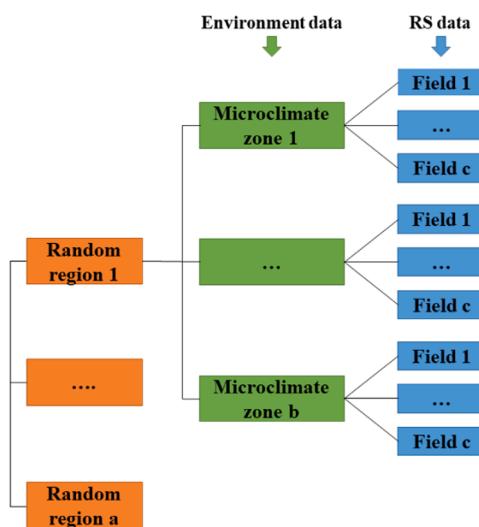


Fig. 9. Nested structure of the remote sensing data and environmental data. The variables labeled as 'a', 'b', and 'c' are the arbitrary numbers with large values.

4.2. Accuracy evaluation of the RHM for the county level data

The wheat yield estimation model in the NCP was constructed by using the multi-source CMID that has not been resampled into multiple time intervals. The validation results of the four methods in four years are listed in Table 4. The accuracy of using full data to estimate yield was significantly better than utilizing only RS data or environmental data. Based on the RS data or environmental data, the accuracy of the three widely used traditional models was unstable in different years of validation. According to the full data, the performance of the developed RHM was slightly better than the widely used traditional model, followed by the RF model. The accuracy of the all models was significantly improved by resampling the data into multiple time intervals. The results are shown in Table 5. Overall, the yield estimation accuracy of all models (except MLR) with full data was better than that using only RS data and environmental data. In comparison with the other models, the RHM with full data was the best model, with an average R^2 of 0.62, MAE of 725.58 kg/ha, RMSE of 877.67 kg/ha, and nRMSE of 12.83%. The second one was the RF model with full data, whose precision was higher than that of SVM and MLR. In comparison with the RF, the estimation error of the RHM was slightly reduced, with an average MAE decrease of 36.44 kg/ha and nRMSE decrease of 0.78%. The accuracy improvement of the RHMs developed based on the CMID was relatively small compared with that of the FMD. Meanwhile, the MLR and HLM suffered from overfitting due to a considerable number of input variables, resulting in poor precision. The Tables A2 and A3 in the appendix provide additional insights into the estimation accuracy of all models based on the COS. This supplementation enhances the comparative analysis between the models utilizing county-level data and those employing FMD. The estimation results of the COS were basically consistent with CMID, and the performance of all models based on the time-phased multisource data was better than that of the models without time-phased data. The RHM with full data was the best COS based on the wheat yield estimation model, with an average R^2 of 0.68, MAE of 519.98 kg/ha, RMSE of 649.15 kg/ha, and an average nRMSE of 11.41%, which was superior to the various models by using other data. The developed RHM has slightly higher accuracy than the second model (i.e., RF), with an average MAE reduction of 86.37 kg/ha and nRMSE reduction of 1.90%.

4.3. Cross-validation of the performance of the models with different spatial scales

Pairwise cross-validation was conducted using FMD, CMID, and COS to better understand the performance differences of the RHM at different spatial scales. For example, the model trained by FMD was validated by using county-level data (e.g., CMID or COS). The cross-validation results of the 1 year leave are shown in Fig. 6. The validation results of the RHM at the same spatial scale were satisfactory, and the accuracy was relatively stable in any of the 4 years. The CMID and COS significantly increased the validation error of the RHM based on the FMD compared with the validation results of FMD (Fig. 6a). Fig. 6b–c shows that the validation of the CMID and COS models on the FMD resulted in significant errors. Significant differences were also observed between the two county-scale models. These results indicate that the three models are not universal, and models built on one spatial scale may produce large errors when applied to another spatial scale.

4.4. Optimal estimation time provided by the RHM

Given the time resampling of the whole growth stages of wheat, the developed RHM could estimate the wheat yield within the lead time of 14 solar terms. The comparison result of the accuracy of the models established at different time indicated that the difference in the optimal estimation time of RHMs at different spatial scales could be determined. Fig. 7 shows the normalized values of the average nRMSE from 0% to 100% to facilitate the disclosure of the magnitude of changes observed for the performance of the different spatial scale models during ST1 to ST14 (method reference Feng et al., 2020a and Li et al., 2021a). The accuracy of the three spatial scale models increased with the time of wheat growth and development, while the rate of increase was relatively low in the early and late stages of growth. The performance improvement of the wheat yield estimation model based on FMD constructed in the ST12 was the greatest, with a 28% increase in the normalized values of nRMSE. After ST12, the model accuracy slightly improved. Accordingly, the ST12 was the most suitable lead time for the wheat yield estimation at the field scale, which was the Beginning of Summer (early May) among the 24 solar terms in China. ST9 to ST12 were the optimal time stages for estimating the wheat yield at the field scale, and the model accuracy has been improved by 63%. The response of the two county-level scale models of CMID and COS to different time intervals was similar. The performance improvement of the wheat yield estimation model based on CMID was the greatest in the ST9 stage, with a 21% increase in the normalized values of nRMSE. The performance improvement of the wheat yield estimation model based on COS was the greatest in the ST10 stage, with a 18% increase in the normalized values of nRMSE. Accordingly, ST9 and ST10 were the most suitable lead times for the CMID and COS wheat yield estimation, respectively. ST8 to ST11 were the optimal time stages for estimating wheat yield at the county-level scale, with improved model accuracy rates of 46% and 36%.

4.5. Parameter importance

Relative importance was used to assess the contribution of the different input variables to yield estimates in the RHM. The variable importance of the models at varying time intervals is shown in Fig. 8. The SR, AT, soil, and vegetation indices are the key variables for the RHM to estimate the wheat yield. The relative importance of the key input variables varied for the RHMs established at different time intervals. The relative importance of the key variables also varies for models at different spatial scales. During ST1 to ST7, SR was the primary variable for RHM constructed by FMD, followed by SWVL. After ST8, STL was the primary estimation variable for the RHM constructed by FMD. During ST1–ST5, SR was the main estimation variable for the RHM constructed by CMID, followed by SWVL. After ST6, the VI was the

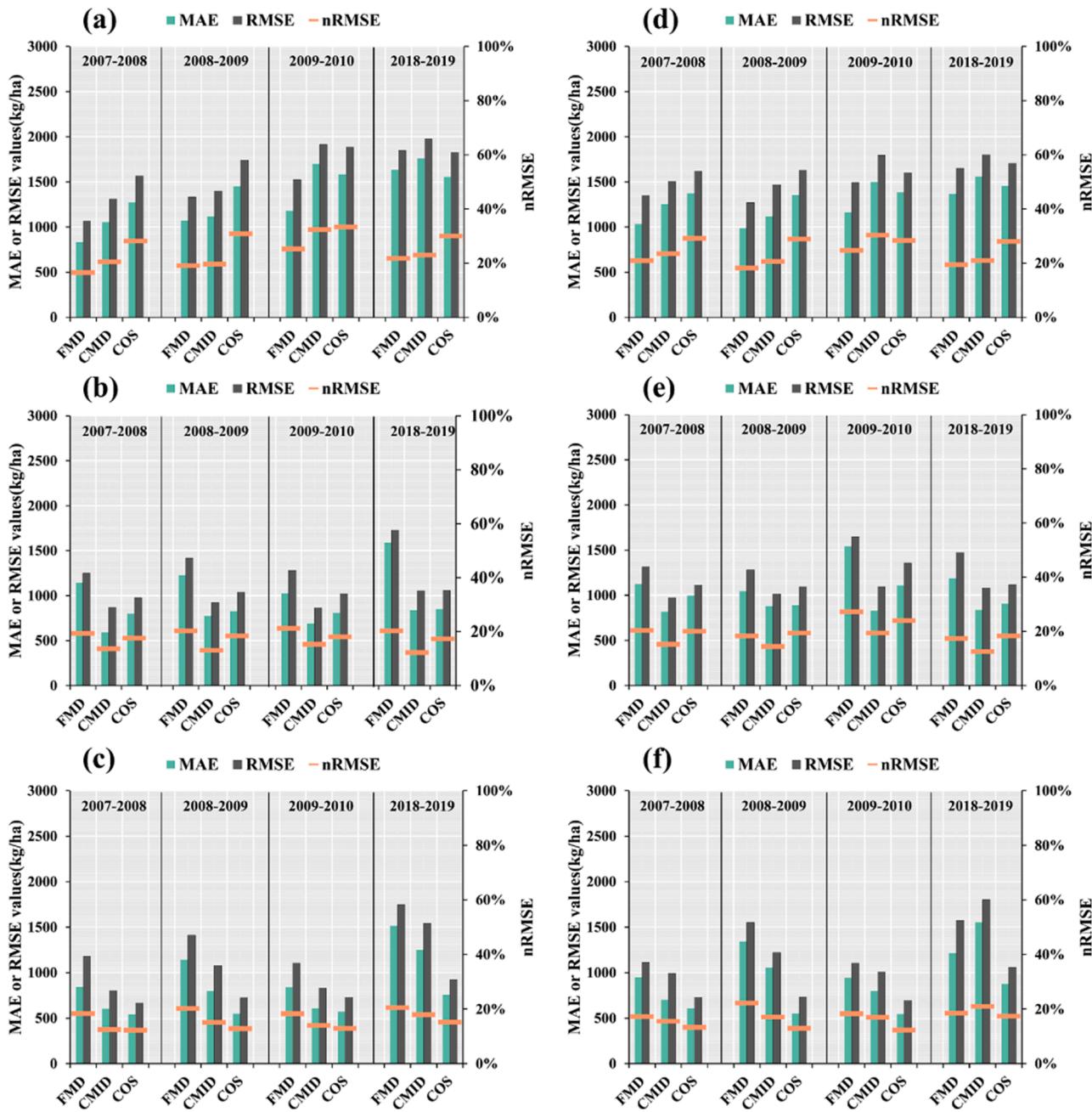


Fig. 10. Comparison of cross validation accuracy of other models with different spatial scales. The models of panels (a)–(c) are RF models, and the models of panel (d)–(f) are the SVM model. Panel (a) and (d) indicates that all three spatial scale models are validated by using the FMD. Panel (b) and (e) indicates that all three spatial scale models are validated by using the CMID. Panel (c) and (f) indicates that all three spatial scale models are validated using the COS. The horizontal coordinates represent the validation data of the model in the corresponding year.

primary estimation variable of the RHM constructed by CMID, with GCI performing the best. AT was the main estimation variable for the RHM constructed by COS in ST1 and ST2. After ST3, VI was the most important estimation variable, with EVI being the optimal. Some key variables also had similar relative importance at different spatial scales (such as SWVL and SR in Fig. 8j–n), while others had significant differences (such as STL and EVI in Fig. 8h–n). In this study, the relative importance of the VIs to county-level models (CMID and COS) was significantly greater than that of the FMD models. The importance of PRE, WS, and VP was not highly ranked for estimating wheat yield using RHM.

5. Discussion

5.1. Performance of the RHM

In this study, the wheat yield estimation system in the NCP was constructed considering data hierarchy and model integration by integrating environmental and RS factors. The wheat yield at different spatial scales (field and county level scales) was estimated using the measured and statistical data from different data sources, and satisfactory results were obtained. Previous studies also used crop models and machine learning models to estimate the wheat in the NCP. Chen et al. (2018) estimated the wheat yield in the NCP based on the crop model (MCWLA-Wheat) and using RS data to assimilate the phenology and leaf

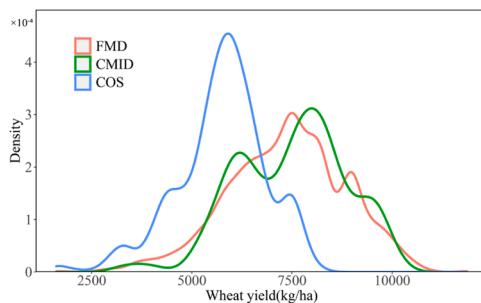


Fig. 11. Probability density curve of wheat yield at different spatial scales in the North China Plain.

area index, with RMSE of 737 kg/ha. The proposed model had a higher yield estimation accuracy with RMSE of 649.15 kg/ha compared with their results. Li et al. (2021a) successfully estimated the wheat yield in China by using the RF and multi-source environmental data, and their reported accuracy was better than that of Silvestro et al. (2017). Tables 2–5 indicate that the developed RHM achieved more reasonable accuracy than the RF model in estimating yield. Therefore, our wheat yield estimation method achieved similar or even better performance in terms of accuracy than most previous studies.

Previous studies estimated wheat by using multi-source meteorological data (Wang et al., 2023), spatial RS data (Ji et al., 2020), and environmental data (Zhou et al., 2022) and achieved satisfactory results by using the powerful nonlinear fitting ability of machine learning. However, these studies rarely focus on the hierarchical nesting relationship and independence between different categories of data, and only consider multiple factors that affect wheat yield as parallel parameters in nonlinear models (i.e., black boxes). The hierarchical nested relationship between RS observations and environmental data is shown in Fig. 9. Given the geographical pattern of climate and the coarse resolution of climate data, multiple heterogeneous RS observations and yield data in a microclimate zone or similar climate zone were obtained (assuming that the environmental variables remain unchanged). RS data can effectively explain the changes in yield within the microclimate zone. However, the environmental changes make the relationship between RS data and yield (intra layer relationship) not unique, and the changes in intra layer relationships in different microclimate zones are regulated by environmental factors (inter layer relationship). This phenomenon of hierarchical nesting is widely present, and the HLMs provide powerful analytical methods, which are frequently used in social work and geoscience (Gavin and Hofmann, 2002; Li et al., 2020; Xu et al., 2020). This study integrated multiple HLMs to construct an RHM. The random effects of the RHM can leverage the relationship between environmental variables to connect multiple intra layer sub models and explain the percentage changes of the environmental variables at the regional and interannual scales. The variance of the model is reduced, and good generalization ability and anti-overfitting ability are achieved due to the randomness of the RHM sample and feature selection.

In addition, previous studies have mostly utilized detailed phenological data (Li et al., 2021a), monthly synthetic data (Zhou et al., 2022), or rough inversion of phenology using RS information (Meroni et al., 2021; Yuchuan Luo et al., 2020). However, it is difficult to accurately obtain large-scale phenological information, which often hinders the promotion and application of the model (Curnel et al., 2011). Monthly synthetic data often ignores spatial heterogeneity (Jin et al., 2018). The phenology of satellite RS inversion is limited by factors such as resolution, sensors, and cloud pollution, greatly increasing the uncertainty of data refinement (Younes et al., 2021). This study subdivided the entire growth stage of wheat according to the 24 solar terms in China. The time interval features selected for the samples in different regions vary, and the sub-model in the RHM that selects the best time interval feature is retained. This condition can further correct the estimation bias caused

by large-scale phenological differences. The 24 solar terms in China are easy to obtain and promote, and more in line with the farming habits of farmers in the NCP. The wheat yield estimation system proposed in this study was a feasible approach at the field and county-level spatial scales, which can be extended to other regions and years to obtain new insights for guiding agricultural production and specifying food strategies.

5.2. Different spatial scales

The cross-validation results between RHM models at different spatial scales (Fig. 6) indicate that the model produces significant errors when applied at different spatial scales. In contrast, paired cross validation was conducted using data from different spatial scales and other models (RF and SVM). The cross-validation results of the 1 year leave are shown in Fig. 10. Although other models also produce errors when applied at different spatial scales, the difference in errors is relatively small. In addition, the RHM at the field scale has a better improvement in the yield estimation accuracy than at the county scale compared with the other models. The reason may be that the average effect and variable heterogeneity within the county are not considered in the county scale model, and the field scale model redistributes the relationship between RS and yield based on the environmental data (rather than administrative counties). This situation is also the reason why the models established at different spatial scales are not universal, even if the FMD and CMID are measured by the same standard. The multi-source data at the field scale have a significant hierarchical structure, while the county data confuse the hierarchical relationship of the multi-source data due to the average effect. Accordingly, the heterogeneity of the RHM at different spatial scales can be well explained. However, other models did not consider the hierarchical relationships between different types of data, and the heterogeneity of models at different spatial scales was not significant compared to RHM. Therefore, agricultural practitioners need to be particularly cautious when constructing RHM using data obtained at different spatial scales.

In addition, the average effect and observation methods result in differences in the optimal estimation lead time and driving factors at different spatial scales. The precision of the wheat yield estimation model began to significantly improve at ST9 (ST8 of the county level scale) because the wheat in most areas of the NCP passed through the winter and began jointing, and the stabilized growth enabled the variables screened by the model to better explain the yield formation (Li et al., 2021a; Otteson, 2007). After ST12, the performance growth rate of the wheat yield estimation model decreased, possibly because the grains of wheat were determined after the filling period (Szczepaniak et al., 2022), and the effect of the RS and environmental factors on the yield estimation was relatively small thereafter. In the early stages (ST1–ST7 of the field scale and ST1–ST5 of the county level), the important influencing variables of the yield estimation model were SR and SWVL (Fig. 8). The main reason may be that the winter wheat in the NCP was in the tillering stage, and the quality of tillering directly affected the wheat yield. Light affected the tillering by acting on the stem base, and sufficient light combined with the sparse seedling density can increase the number of tillers per plant (Ochagavía et al., 2017; Shang et al., 2021). Water shortage is a known limiting factor for the sustainable production of wheat in the NCP (Kan et al., 2020). The soil moisture and drying conditions frequently influence the potential tillering ability of wheat by affecting the coleoptile of wheat seed germination and first tiller (Blum et al., 1990). After ST8, STL was the most important estimation variable at the field scale. The soil temperature affects the absorption of nutrients and water, the root growth during the vegetative growth stage, and the number of wheat grains and grain filling during the reproductive growth stage (Li et al., 2021b). The relative importance of the VI after ST6 of the CMID (ST3 of COS) is significantly higher than that of the environmental variables. The reason may be that the average effect of multiple pixels enhances the expression ability of the VI on the overall growth of wheat in the county, which is

similar to the study by Jiang et al. (2020).

5.3. Limitations

This study established an RHM yield estimation model and estimated the wheat yield in the NCP at different spatial scales, but few limitations persisted. To better compare the differences between different spatial scale models, the COS data was used in this study, but it is evident that the method of obtaining this data is vastly different from CMID and FMD. Fig. 11 shows the density maps of raw data at three spatial scales. It can be observed that the data ranges of CMID and FMD are similar but different, while the data ranges of COS and other data have significant differences but similar trends. The difference between COS and other spatial scale data is likely due to different observation methods leading to inconsistent data ranges, while the difference between CMID and FMD is more due to average effects. In addition, this study used 4 years of data to construct RHM and achieved satisfactory results, and each year's data was independently validated. However, due to experimental limitations, the 4-year data is not continuous, and further continuous experimental data needs to be obtained. While the model takes into account climate, soil, and growth information, the impact of latitude variations, diverse local varieties, and field management practices (e.g., irrigation and fertilization) on yields cannot be disregarded. Future research needs to transcend limitations related to a broad scope and observational conditions, delving further into the adaptability of the model to additional influencing factors. In addition, the selection of time intervals was empirical and random, and further work can utilize environmental data to digitize phenological information (Li et al., 2022b).

6. Conclusion

In this study, an RHM was constructed by integrating multiple HLMs with multi-source environment and RS variables, and the heterogeneity of the model was evaluated at different spatial scales. First, this study refined the multi-source data of the whole growth stage of wheat in the NCP based on the 24 solar terms in China, providing more time options for yield estimation. Second, a dynamic wheat yield estimation model based on the RHM was developed by integrating HLMs and ensemble learning techniques. Finally, model heterogeneity across spatial scales

was explored, considering hierarchical differences in data, optimal estimation time, and parameter importance. The results indicated that the satisfactory accuracy of the RHM was achieved on FMD ($R^2 = 0.52$, MAE = 908.87 kg/ha, nRMSE = 16.43%), CMID ($R^2 = 0.62$, MAE = 725.58 kg/ha, nRMSE = 12.83%), and COS ($R^2 = 0.68$, MAE = 519.98 kg/ha, nRMSE = 11.41%) and significantly superior to the widely used traditional models. Significant heterogeneity and non-universality were observed in the yield estimation models at different spatial scales, and the RHM performed better at the field scale. The mining of hierarchical and interactive relationships between multi-source data compensated for the blind and unknowable shortcomings of the traditional machine learning models in feature utilization and improved the accuracy of large-scale and interannual yield estimation. Future work will validate the applicability of the RHM to other regions and crops across the country.

CRediT authorship contribution statement

Xiaobin Xu: Conceptualization, Data curation, Methodology, Validation, Visualization, Writing – original draft. **Wei He:** Resources, Supervision, Validation, Writing – review & editing. **Hongyan Zhang:** Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Data Availability

The authors do not have permission to share data.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3903605 and National Natural Science Foundation of China under Grant 42071322.

Appendix

Table A1. Data statistics used in this study.

Category	Year	Min	Max	Mean	Median	SD	CV	Sample size
FMD	2007–2008	3519.00	8935.29	6432.22	6450.00	997.56	999.44	270
	2008–2009	1720.09	11868.09	6968.51	7035.72	1204.65	1206.74	290
	2009–2010	2640.13	9224.06	6027.90	6282.31	1355.33	1360.00	140
	2018–2019	4000.00	10875.00	8498.55	8250.00	930.39	931.32	500
CMID	2007–2008	5307.00	8238.36	6368.24	6233.60	685.66	701.07	21
	2008–2009	5155.15	8322.27	7033.58	7091.97	829.67	849.20	20
	2009–2010	3366.66	7354.39	5629.63	5999.39	1194.31	1266.76	8
	2018–2019	7545.00	10025.00	8588.14	8350.45	716.07	723.65	38
COS	2007–2008	1645.07	6781.00	5463.66	5805.95	1226.22	1245.84	32
	2008–2009	1714.72	7710.14	5627.84	5745.01	972.79	976.17	40
	2009–2010	1616.16	7533.00	5635.74	5735.50	947.23	950.32	25
	2018–2019	3315.08	8212.27	6067.36	6291.75	1311.34	1316.22	68

Note: the numerical unit presented in the table is kilograms per hectare (kg/ha); SD denotes the standard deviation; CV denotes the coefficient of variation. Table A2. Performance evaluation of the four models constructed by using the non-time-phased multisource data and county-level official statistics (COS) during the wheat growth stages.

Year	Error (kg/ha)	RS data			Environmental data			Full data				
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	HLM	RHM
2007–2008	R^2	0.51	0.49	0.53	0.58	0.57	0.62	0.61	0.61	0.65	0.63	0.66
	MAE	884.35	1035.89	850.40	755.15	736.36	607.55	655.88	697.81	547.59	605.25	516.71
	RMSE	1245.55	1275.86	1118.64	1014.45	1061.92	787.21	836.97	826.65	726.56	792.39	680.91

(continued on next page)

(continued)

Year	Error (kg/ha)	RS data			Environmental data			Full data					
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	HLM	RHM	
2008–2009	nRMSE	22.80%	23.35%	20.47%	18.57%	19.44%	14.41%	15.32%	15.13%	13.30%	14.50%	12.46%	
	R ²	0.62	0.64	0.62	0.65	0.63	0.64	0.67	0.66	0.66	0.67	0.67	
	MAE	818.91	713.40	791.60	630.19	762.50	677.17	586.97	624.55	607.37	585.15	568.43	
	RMSE	1011.97	895.09	963.72	819.19	906.51	867.50	727.06	766.12	757.04	726.01	715.88	
2009–2010	nRMSE	17.98%	15.90%	17.12%	14.56%	16.11%	15.41%	12.92%	13.61%	13.45%	12.90%	12.72%	
	R ²	0.48	0.48	0.48	0.50	0.59	0.50	0.60	0.60	0.61	0.60	0.62	
	MAE	1091.76	1062.19	1062.33	816.21	726.77	902.90	673.74	676.26	615.35	658.02	567.61	
	RMSE	1290.53	1268.07	1267.91	1080.57	878.52	1117.17	805.51	815.29	774.60	789.53	703.31	
2018–2019	nRMSE	22.90%	22.50%	22.50%	19.17%	15.59%	19.82%	14.29%	14.47%	13.74%	14.01%	12.48%	
	R ²	0.63	0.63	0.63	0.58	0.58	0.55	0.60	0.55	0.63	0.62	0.65	
	MAE	718.32	754.16	770.36	859.61	891.04	1000.91	815.40	956.25	787.15	808.24	700.00	
	RMSE	946.07	958.96	944.88	1083.30	1058.61	1187.47	1005.43	1164.53	947.16	970.84	856.16	
	nRMSE	15.59%	15.81%	15.57%	17.85%	17.45%	19.57%	16.57%	19.19%	15.61%	16.00%	14.11%	

Table A3. Performance evaluation of the four models constructed by using the time-phased multisource data and COS during the wheat growth stages.

Year	Error (kg/ha)	RS data			Environmental data			Full data					
		MLR	SVM	RF	MLR	SVM	RF	MLR	SVM	RF	HLM	RHM	
2007–2008	R ²	0.40	0.65	0.66	0.04	0.65	0.66	0.03	0.65	0.66	0.05	0.68	
	MAE	1020.28	572.34	545.46	2185.06	602.18	512.97	2163.70	610.34	544.76	1854.96	509.32	
	RMSE	1438.69	716.56	690.83	2583.82	728.74	670.31	3087.54	728.94	668.26	2737.24	650.71	
	nRMSE	26.33%	13.11%	12.64%	47.29%	13.34%	12.27%	56.51%	13.34%	12.23%	50.10%	11.91%	
2008–2009	R ²	0.65	0.65	0.63	0.00	0.64	0.64	0.00	0.65	0.67	0.02	0.70	
	MAE	581.38	603.37	684.49	3861.76	651.65	616.42	2402.70	552.15	550.96	2300.11	470.75	
	RMSE	773.92	767.44	862.77	4922.77	813.97	838.77	3420.63	732.23	723.71	3313.63	610.51	
	nRMSE	13.75%	13.64%	15.33%	87.47%	14.46%	14.90%	60.78%	13.01%	12.86%	58.89%	10.85%	
2009–2010	R ²	0.05	0.44	0.48	0.00	0.59	0.60	0.00	0.63	0.62	0.03	0.63	
	MAE	1985.46	1112.92	1069.33	4741.97	747.18	688.72	9580.50	550.20	571.48	8102.03	565.06	
	RMSE	2608.37	1369.05	1291.19	5531.21	870.38	812.59	12562.87	695.65	728.54	10442.52	684.33	
	nRMSE	46.28%	24.29%	22.91%	98.15%	15.44%	14.42%	222.91%	12.34%	12.93%	185.25%	12.14%	
2018–2019	R ²	0.47	0.47	0.51	0.00	0.62	0.47	0.00	0.60	0.62	0.03	0.69	
	MAE	1089.69	1129.83	1022.00	4046.57	828.49	1161.88	3301.84	880.07	758.19	3017.76	534.80	
	RMSE	1385.13	1352.23	1245.82	4724.50	990.04	1393.37	3824.76	1057.82	922.63	3592.66	651.06	
	nRMSE	22.83%	22.29%	20.53%	77.87%	16.32%	22.96%	63.04%	17.43%	15.21%	59.21%	10.73%	

References

- Alexandrov, V.A., Hoogenboom, G., 2000. The impact of climate variability and change on crop yield in Bulgaria. *Agric. For. Meteorol.* 104 (4), 315–327.
- Azzari, G., et al., 2019. Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. *Remote Sens. Environ.* 221, 417–429.
- Basso, B., Liu, L., 2019. Seasonal crop yield forecast: methods, applications, and accuracies. In: Sparks, D.L. (Ed.), *Advances in Agronomy*. Academic Press, pp. 201–255.
- Bergstra, J. and Bengio, Y.J., 2012. Random search for hyper-parameter optimization. *13(2)*.
- Blum, A., Ramaiah, S., Kanemasu, E.T., Paulsen, G.M., 1990. Wheat recovery from drought stress at the tillering stage of development. *Field Crops Res.* 24 (1), 67–85.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Buontempo, C., et al., 2020. Fostering the development of climate services through Copernicus Climate Change Service (C3S) for agriculture applications. *Weather Clim. Extrem.* 27, 100226.
- Burke, M. and Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *114(9)*: 2189–2194.
- Cai, Y., et al., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* 274, 144–159.
- Cao, H.-z., et al., 2019. Identifying the limiting factors driving the winter wheat yield gap on smallholder farms by agronomic diagnosis in North China Plain. *J. Integr. Agric.* 18 (8), 1701–1713.
- Cao, J., et al., 2021. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *Eur. J. Agron.* 123, 126204.
- Chen, Y., Zhang, Z., Tao, F., 2018. Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. *Eur. J. Agron.* 101, 163–173.
- Cheng, M., et al., 2022. Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China. *Agric. For. Meteorol.* 323.
- Curnel, Y., de Wit, A.J.W., Duveiller, G., Defourny, P., 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agric. For. Meteorol.* 151 (12), 1843–1855.
- Dong, J., et al., 2020. Early-season mapping of winter wheat in China based on Landsat and Sentinel images. *Earth Syst. Sci. Data* 12 (4), 3081–3095.
- Duan, J., et al., 2019. Optimizing nitrogen management to achieve high yield, high nitrogen efficiency and low nitrogen emission in winter wheat. *Sci. Total Environ.* 697, 134088.
- Feng, C., et al., 2020a. A method for prediction of thermophilic protein based on reduced amino acids and mixed features. *Front. Bioeng. Biotechnol.* 8.
- Feng, P., et al., 2020b. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agric. For. Meteorol.* 285–286.
- Franch, B., et al., 2015. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. *Remote Sens. Environ.* 161, 131–148.
- Fu, Z., et al., 2022. Combining UAV multispectral imagery and ecological factors to estimate leaf nitrogen and grain protein content of wheat. *Eur. J. Agron.* 132, 126405.
- Gao, B.-C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58 (3), 257–266.
- Gavin, M.B., Hofmann, D.A., 2002. Using hierarchical linear modeling to investigate the moderating influence of leadership climate. *Leadersh. Q.* 13 (1), 15–33.
- Gitelson, A.A., Viña, A., Ciganda, V., Rundquist, D.C. and Arkebauer, T.J.J.Grl, 2005. Remote estimation of canopy chlorophyll content in crops. *32(8)*.
- Huang, J., et al., 2019. Assimilation of remote sensing into crop growth models: current status and perspectives. *Agric. For. Meteorol.* 276–277, 107609.
- Huete, A., et al., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83 (1), 195–213.
- Ji, Z., Pan, Y., Li, M., 2020. Using NDVI Time Series Curve Change Rate to Estimate Winter Wheat Yield. *IGARSS 2020 - 2020 IEEE Int. Geosci. Remote Sens. Symp.* 5183–5186.
- Jiang, H., et al., 2020. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. *Glob. Chang Biol.* 26 (3), 1754–1766.
- Jiang, Z., Huete, A., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 112 (10), 3833–3845.
- Jin, X., et al., 2018. A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* 92, 141–152.
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agric. Water Manag.* 227, 105846.

- Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* 141, 116–128.
- Kan, Z.-R., et al., 2020. Responses of grain yield and water use efficiency of winter wheat to tillage in the North China Plain. *Field Crops Res.* 249, 107760.
- Kang, Y., et al., 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environ. Res. Lett.* 15 (6).
- Kaul, M., Hill, R.L., Walther, C., 2005. Artificial neural networks for corn and soybean yield prediction. *Agric. Syst.* 85 (1), 1–18.
- Kogan, F., et al., 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *Int. J. Appl. Earth Obs. Geoinf.* 23, 192–203.
- Launay, M., Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agric., Ecosyst. Environ.* 111 (1), 321–339.
- Lecerf, R., Ceglar, A., López-Lozano, R., Van Der Velde, M., Baruth, B., 2019. Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. *Agric. Syst.* 168, 191–202.
- Leghari, S.J., Hu, K., Liang, H., Wei, Y., 2019. Modeling water and nitrogen balance of different cropping systems in the North China Plain. *Agronomy* 9 (11), 696.
- Leroux, L., et al., 2019. Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. *Eur. J. Agron.* 108, 11–26.
- Li, L., et al., 2021a. Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. *Agric. For. Meteorol.* 308–309.
- Li, Y., et al., 2021b. Effects of mulching on soil temperature and yield of winter wheat in the semiarid rainfed area. *Field Crops Res.* 271.
- Li, Z., et al., 2015. Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *Eur. J. Agron.* 71, 53–62.
- Li, Z., et al., 2020. A hierarchical interannual wheat yield and grain protein prediction model using spectral vegetative indices and meteorological data. *Field Crops Res.* 248.
- Li, Z., et al., 2022b. Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sens. Environ.* 273.
- Li, Z., Ding, L., Xu, D., 2022a. Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. *Sci. Total Environ.* 815, 152880.
- Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. *Sensors* 18 (8).
- Lobell, D.B., Field, C.B., 2007. Global scale climate–crop yield relationships and the impacts of recent warming. *Environ. Res. Lett.* 2 (1).
- Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agric., Ecosyst. Environ.* 94 (2), 205–220.
- Ma, Y., Zhang, Z., Kang, Y., Özdogan, M., 2021. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* 259.
- Mathieu, J.A., Aires, F., 2018. Assessment of the agro-climatic indices to improve crop yield forecasting. *Agric. For. Meteorol.* 253–254, 15–30.
- Mehdaoui, R., Anane, M., 2020. Exploitation of the red-edge bands of Sentinel 2 to improve the estimation of durum wheat yield in Grombalia region (Northeastern Tunisia). *Int. J. Remote Sens.* 41 (23), 8986–9008.
- Meroni, M., et al., 2021. Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2. *Remote Sens Environ.* 253, 112232.
- Ochagavía, H., Prieto, P., Savin, R., Griffiths, S., Slafer, G.A., 2017. Duration of developmental phases, and dynamics of leaf appearance and tillering, as affected by source and doses of photoperiod insensitivity alleles in wheat under field conditions. *Field Crops Res.* 214, 45–55.
- Otteson, B.N., 2007. Response of Spring Wheat to Varying Seeding Rate and Nitrogen Management.
- Raudenbush, S., Bryk, A., 2001. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, Inc.
- Saad El Imani, H., El Harti, A., El Iysaouy, L., 2022. Wheat yield estimation using remote sensing indices derived from sentinel-2 time series and google earth engine in a highly fragmented and heterogeneous agricultural region. *Agronomy* 12 (11).
- Shang, Q., et al., 2021. Genetic, hormonal, and environmental control of tillering in wheat. *Crop J.* 9 (5), 986–991.
- Silvestro, P., et al., 2017. Estimating wheat yield in china at the field and district scale from the assimilation of satellite data into the aquacrop and simple algorithm for yield (SAFY) models. *Remote Sens.* 9 (5).
- Skakun, S., et al., 2018. Winter wheat yield assessment using landsat 8 and sentinel-2 data. *IGARSS 2018 - 2018 IEEE Int. Geosci. Remote Sens. Symp.* 5964–5967.
- Szczepaniak, W., Grzebisz, W., Potarzyczyk, J., 2022. Yield predictive worth of pre-flowering and post-flowering indicators of nitrogen economy in high yielding winter wheat. *Agronomy* 13 (1).
- Taiwo, B.E., et al., 2023. Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques. *Environ. Sustain. Indic.* 18.
- Tian, H., et al., 2021. An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agric. For. Meteorol.* 310.
- Wang, H., et al., 2016. Monitoring winter wheat drought threat in Northern China using multiple climate-based drought indices and soil moisture during 2000–2013. *Agric. For. Meteorol.* 228–229, 1–12.
- Wang, J., et al., 2020a. Integrating remote sensing-based process model with environmental zonation scheme to estimate rice yield gap in Northeast China. *Field Crops Res.* 246, 107682.
- Wang, Y., Zhang, Z., Feng, L., Du, Q., Runge, T., 2020b. Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous United States. *Remote Sens.* 12 (8).
- Wang, Y., Shi, W., Wen, T., 2023. Prediction of winter wheat yield and dry matter in North China Plain using machine learning algorithms for optimal water and nitrogen application. *Agric. Water Manag.* 277.
- Wu, C., Niu, Z., Gao, S., 2012. The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland. *Ecol. Indic.* 14 (1), 66–73.
- Wu, D., et al., 2021. A new agricultural drought index for monitoring the water stress of winter wheat. *Agric. Water Manag.* 244.
- Xiao, C., et al., 2019. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sens. Environ.* 233, 111358.
- Xu, X., et al., 2020. Prediction of wheat grain protein by coupling multisource remote sensing imagery and ECMWF data. *Remote Sens.* 12 (8).
- Xu, X., et al., 2021. A comprehensive yield evaluation indicator based on an improved fuzzy comprehensive evaluation method and hyperspectral data. *Field Crops Res.* 270.
- Younes, N., Joyce, K.E., Maier, S.W., 2021. All models of satellite-derived phenology are wrong, but some are useful: a case study from northern Australia. *Int. J. Appl. Earth Obs. Geoinf.* 97.
- Yuchuan Luo, Z.Z., Chen, Yi, Li, Ziyue, Tao, Fulu, 2020. ChinaCropPhen1km: a high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products. *Earth Syst. Sci. Data.*
- Zhang, S., et al., 2023. Integrating a novel irrigation approximation method with a process-based remote sensing model to estimate multi-years winter wheat yield over the North China Plain. *J. Integr. Agric.*
- Zhao, J., et al., 2023. Summer maize growth estimation based on near-surface multi-source data. *Agronomy* 13 (2).
- Zhao, Y., et al., 2022. Drought monitoring and performance evaluation based on machine learning fusion of multi-source remote sensing drought factors. *Remote Sens.* 14 (24).
- Zhou, Q., et al., 2023. Winter crop rotation intensification to increase rice yield, soil carbon, and microbial diversity. *Heliyon* 9 (1), e12903.
- Zhou, W., et al., 2022. Integrating climate and satellite remote sensing data for predicting county-level wheat yield in China using machine learning methods. *Int. J. Appl. Earth Obs. Geoinf.* 111.