



Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Accurately mapping global wheat production system using deep learning algorithms



Yuchuan Luo ^a, Zhao Zhang ^{a,*}, Juan Cao ^a, Liangliang Zhang ^a, Jing Zhang ^a, Jichong Han ^a, Huimin Zhuang ^a, Fei Cheng ^a, Fulu Tao ^{b,c,d}

^a Academy of Disaster Reduction and Emergency Management Minsitry of Emergency Management & Ministry of Education, School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, China

^b Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

^c College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

^d Natural Resources Institute Finland (Luke), FI-00790 Helsinki, Finland

ARTICLE INFO

Keywords:

Wheat
Crop mapping
Yield estimation
Deep learning
Remote sensing

ABSTRACT

Assessing global food security and developing sustainable production systems need spatially explicit information on crop harvesting areas and yields; however the available datasets are spatially and temporally coarse. Here, we developed a general framework, Global Wheat Production Mapping System (GWPMS), to map the spatial distribution of wheat harvesting area and estimate yield using data-driven models across eight major wheat-producing countries worldwide. We found GWPMS could not only generate robust wheat maps with R^2 consistently greater than 0.8, but also successfully captured a substantial fraction of yield variations with an average of 76%. The developed long short-term memory model outperformed other machine learning algorithms because it characterized the nonlinear and cumulative impacts of meteorological factors on yield. Using the derived wheat maps improved R^2 by 6.7% compared to a popularly used dataset. GWPMS is able to map spatial distribution of harvesting areas in a scalable way and further estimate gridded-yield robustly, and it can be applied globally using publicly available data. GWPMS and the resultant datasets will greatly accelerate our understanding and studies on global food security.

1. Introduction

Wheat provides nearly 20% of the total dietary calorie and protein requirements for 4.5 billion people, with a total harvesting area of 215.9 Mha (FAO, 2019). With the unprecedented growth of world population towards a likely 9 billion by 2050, wheat production necessitates doubling to feed that demand (Nelson et al., 2010). However, in recent years, 37% of global wheat harvesting areas have experienced yield stagnation, challenging the ability of agricultural production to content with the rising global demands (Ray et al., 2012). Thus, timely and accurately mapping wheat production over large spatial extents has important implications for global food security warning, agricultural land use optimizing and decision-making (Nelson and Burchfield, 2021).

Numerous studies have conducted crop yield prediction using satellite-based indices, climatic variables and soil properties (Cai et al., 2019; Cao et al., 2021; Jeong et al., 2022). Remote sensing approaches

often depend on calibrated relationships between vegetation indices (VIs) and crop yields (Lai et al., 2018; Peng et al., 2020); however, VIs provide indirect measurements of yield (Rembold et al., 2013). Climate variables enable directly capturing environmental information on crop growth status, rather than most satellite-based indices (Cai et al., 2019; Luo et al., 2021). Soil properties, as static parameters compared with the above transient ones, also provide additional information to improve crop yield estimation due to their large heterogeneity (Folberth et al., 2016). With more data available at higher resolutions both in space and time, such multi-sources information will more hugely contribute to estimating yields worldwide.

Another improvement road is potentially come from algorithms, especially for machine learning (ML) and deep learning (DL). Statistical models or/and process-based crop models remain powerful and prevalent tools for yield estimation (Deines et al., 2021; Li et al., 2019). Each of them has pros and cons. Statistical models are usually applied over

* Corresponding author.

E-mail address: zhangzhao@bnu.edu.cn (Z. Zhang).

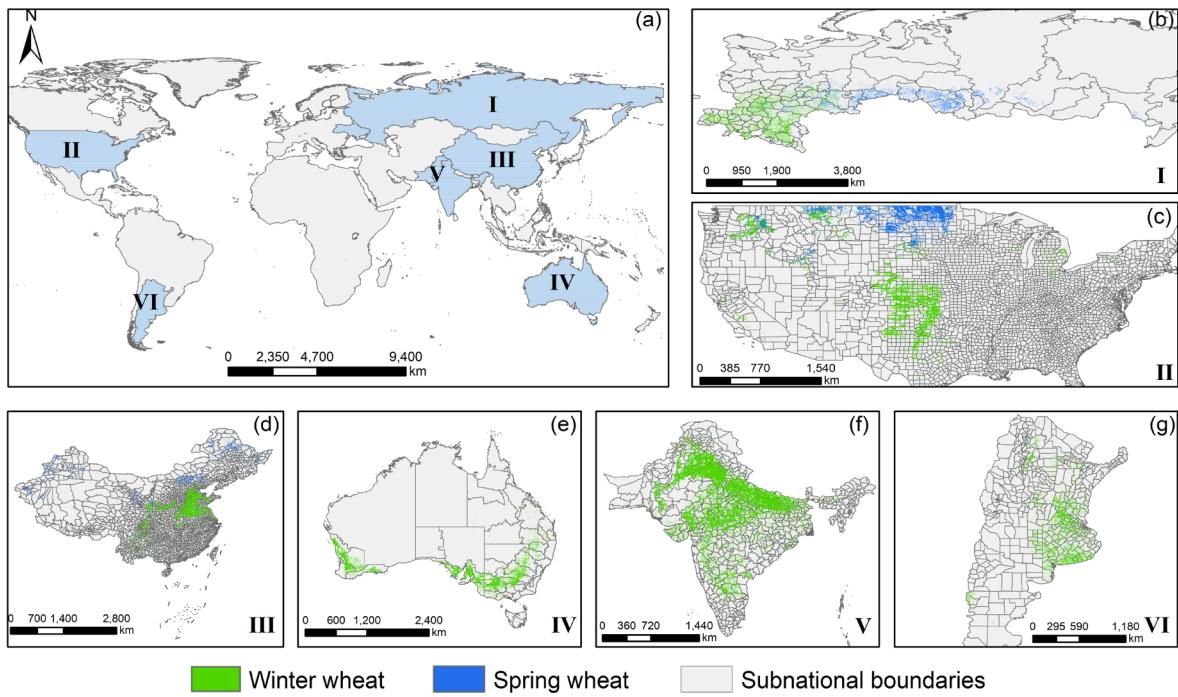


Fig. 1. Location of the study areas covering eight countries globally (a) and spatial distribution of spring and winter wheat in Ukraine (UKR) and Russian Federation (RUS) (b), the United States (USA) (c), China (CHN) (d), Australia (AUS) (e), India (IND) and Pakistan (PAK) (f), and Argentina (ARG) (g).

larger spatial extents with the specific advantages of requiring fewer inputs and having higher computational efficiency (de Wit and van Diepen, 2008; Lobell and Burke, 2010). Compared to traditional linear regression models, ML and DL methods have advantages of learning nonlinear relationships between high dimensional inputs and yield and being immune to multicollinearity, showing a better performance for yield estimation (Sakamoto, 2020). As a state-of-the-art ML algorithm, DL can capture more complex nonlinear relationships with multilayer stack that can transform inputs to augment the complexity and abstractness of the representations (LeCun et al., 2015). For example, Long Short-Term Memory (LSTM) networks performed better than Random Forest (RF) in predicting maize yields in the United States (Jiang et al., 2020).

Despite many researchers having hugely contributed to improving yield estimates from data and algorithms, several limitations existed in previous studies. First, most studies used only a constant land cover or crop map instead of accurate or even annually dynamic crop distribution maps (Iizumi et al., 2014b; Schwalbert et al., 2020). The spatially explicit and annual harvesting areas play paramount roles in estimating yield as they are dynamically changed by many factors. Consequently, the dynamic yields will be misestimated by aggregating information based on one constant crop map to capture crop growing conditions over the corresponding statistical unit. Second, previous efforts to estimate crop yield based on ML or DL models were mainly conducted within a nation, or even at a sub-national scale (Cai et al., 2019; Gomez et al., 2021). Mapping spatially explicit crop production for the global major breadbaskets is of great significance, especially for assessing climate change impact and planning effective adaptations, but has not been conducted so far (Iizumi et al., 2014a).

Most extant approaches to accurately map crop yields necessitate field-surveyed yield data that are extremely time- and money-consuming to collect across the globe. Moreover, the empirical regression models built by remote sensing approaches are typically region-specific and require new ground observations for new applications. Recently, the Scalable Crop Yield Mapper (SCYM) was proposed to estimate gridded yield by integrating large sets of scenario simulations derived from crop model to statistical models that link crop yield with

different combinations of remote sensing images without the need for ground observations (Lobell et al., 2015). SCYM has been suggested to be a powerful tool in several countries, which can explain 75% yield variability in the USA while < 50% variability in South Asia and Africa (Jin et al., 2019). Moreover, applying SCYM into new regions strongly depends on the credible simulations from region-specific crop models that are closely associated with the quality of input data (unavailable nowadays) (Azzari et al., 2017). On the other side, a few studies using agricultural statistical data (e.g., Spatial Production Allocation Model, SPAM) are more viable to estimate global gridded-yields (You et al., 2014). However, only three years' data were generated (i.e., 2000, 2005, and 2010), which was too short to depict temporal dynamics in global yields. Finally, till today, few studies have yet systematically investigated the effectiveness and robustness of the ML and DL models in estimating gridded yields across diverse cropping systems in the world. Therefore, it is essential to develop a systematic and automatic framework for accurately mapping global crop production information based on data-driven models, stepwise from harvesting area to yield.

In this study, we introduce a generalized approach to map global wheat production, Global Wheat Production Mapping System (GWPMs), using more elaborate subnational statistics with a larger number of administrative units and a longer time period than existing work (Cao et al., 2021; Iizumi et al., 2014b). The main objectives are: 1) to develop an automatic phenology-based method to map the spatial distribution of wheat; 2) to quantitatively investigate the predictive performance of RF, Light Gradient Boosting Machine (LightGBM) and LSTM algorithms in estimating gridded yields; and 3) to provide an efficient and scalable framework for estimating global gridded wheat yields using optimal data-driven models, based on accurate wheat maps.

2. Materials and methods

2.1. Study area

Eight countries, including the United States, Russian Federation, China, Australia, Argentina, India, Pakistan, and Ukraine, are covered by the study area (Fig. 1). We chose these countries because of their

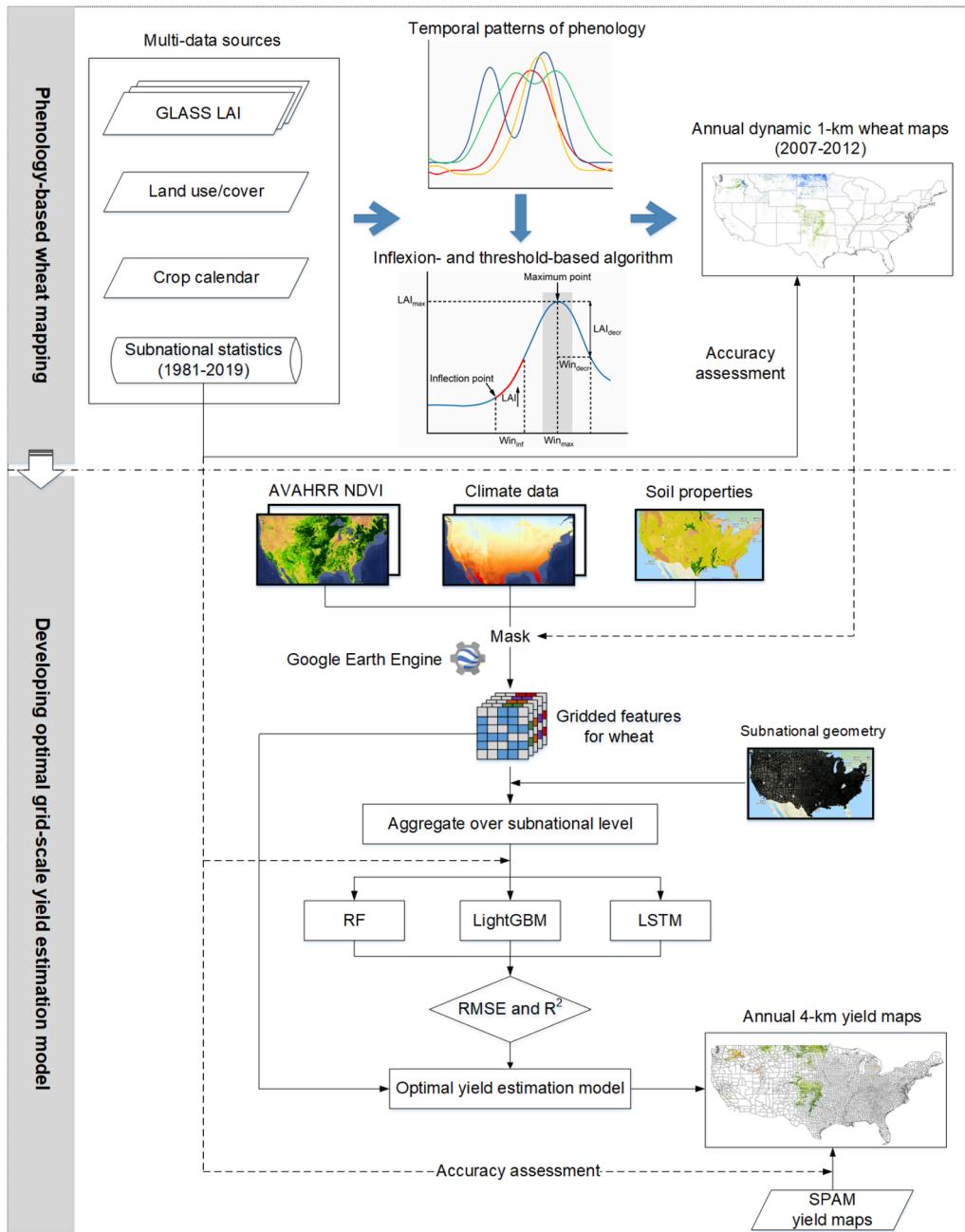


Fig. 2. Illustration of the Global Wheat Production Mapping System (GWPMS).

leading roles in global wheat production and abundant agricultural statistical data, covering ~59% of the global wheat harvesting areas and production (FAO, 2019). In addition, the highly heterogeneous growing environment across the studied countries could examine the scalability and robustness of our methodology. Details for each country were given in Table S1.

2.2. Data

2.2.1. Crop area and yield

Subnational statistics for 6,124 fixed political units across the eight countries were collected, including area harvested (unit: ha) and production (unit: ton). Yield (unit: kg/ha) is defined as the ratio of production and harvesting area. The spatial resolution and time frame of the census data varied across countries (Table S1). Generally, 97.5% of the data were for the administrative unit level 2 or 3 (ADM2 or ADM3) with

the longest time frame spanning from 1981 to 2019. We detected and filtered out the outliers of statistical data with values ranging outside of $\text{mean} \pm 2$ standard deviations. In addition, the annual spatial distributions of wheat areas in China from 2000 to 2015 were obtained from a 1-km dataset named ChinaCropArea1km (Luo et al., 2020).

2.2.2. Satellite data

The satellite data contained daily 0.05° Advanced Very High-Resolution Radiometer (AVHRR) Normalized Difference Vegetation Index (NDVI) from 1981 to 2019, 1-km Global Land Surface Satellite (GLASS) Leaf Area Index (LAI) of 8-day composite from 2006 to 2012 and Global Food Security-support Analysis Data (GFSAD) 1 km cropland data product for the nominal year 2010. AVHRR NDVI was selected due to its specific advantage of the longest records of vegetation growth information compared to other sensors. The GLASS LAI products presented a higher spatiotemporal continuity than other LAI products (<https://www.giss.nasa.gov/data/glass/>)

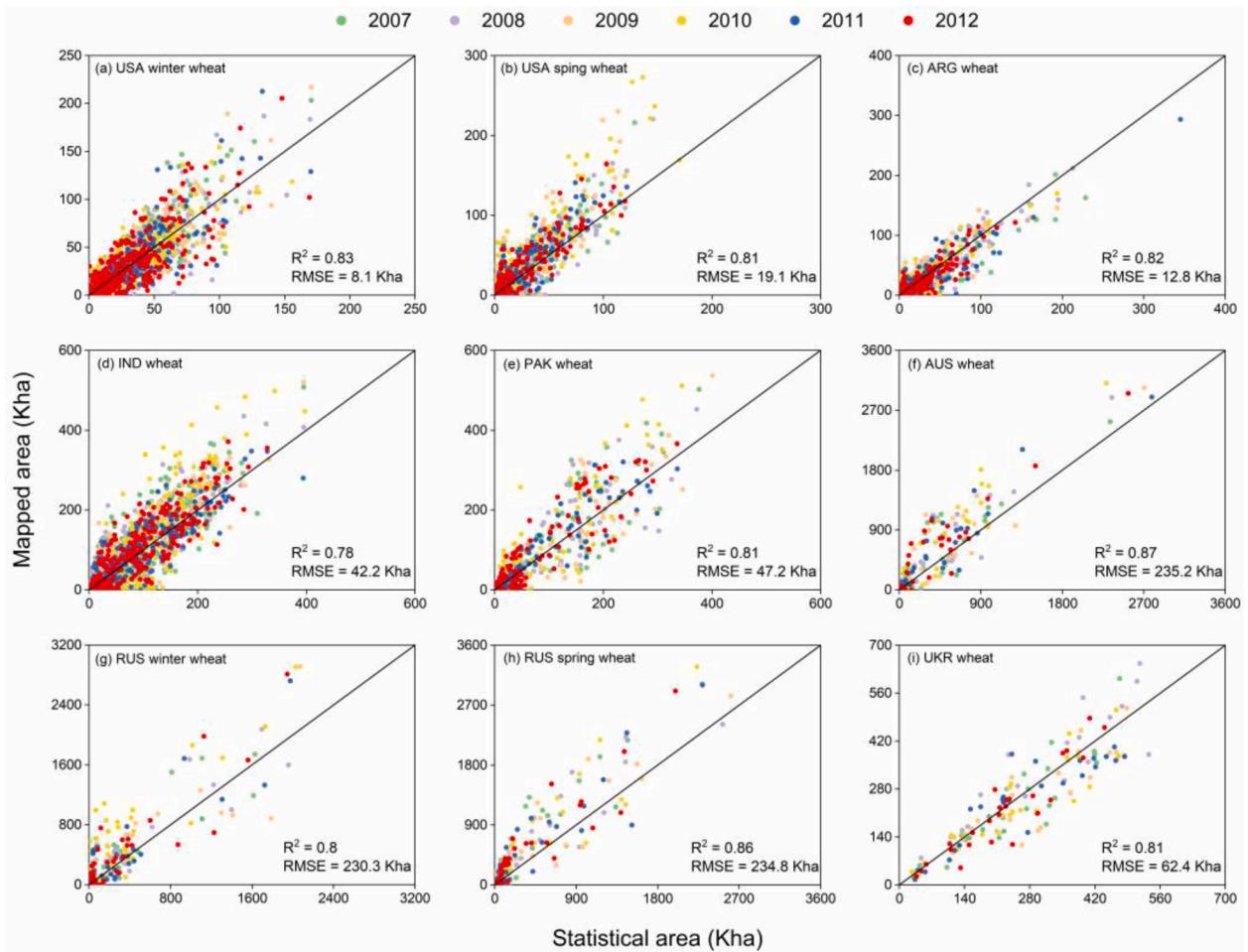


Fig. 3. Subnational-level comparisons between the GWPMS-based wheat harvesting areas and the agricultural census data during 2007–2012.

([://glass-product.bnu.edu.cn/introduction/LAI.html](http://glass-product.bnu.edu.cn/introduction/LAI.html)), having been successfully applied to land cover mapping and phenology monitoring (Xiao et al., 2016; Liu et al., 2020; Luo et al., 2020). GFSAD1km was created using four existing global cropland products with an overall accuracy ranging from 73.3% to 85.2%, including five irrigated and rainfed classes (Yadav and Congalton, 2018).

2.2.3. Climate data and soil properties

Monthly climatic variables were obtained from 4-km TerraClimate datasets (Abatzoglou et al., 2018), including minimum temperatures (Tmax), maximum temperature (Tmin), palmer drought severity index (Pdsi), precipitation (Pre), vapor pressure (Vap), soil moisture (Soil), downward surface shortwave radiation (Srad), evapotranspiration (Pet) and vapor pressure deficit (Vpd) from 1981 to 2019. Soil properties including topsoil (0–30 cm) organic carbon, pH, bulk density, gravel, clay, silt and sand fraction were acquired from Harmonized World Soil Database (HWSD) with a 30 arc-second spatial resolution (FAO, 2012).

2.3. Methods

The flowchart for GWPMS is presented schematically in Fig. 2, consisting of two main components: mapping wheat areas based on phenology and developing optimal grid-scale yield estimation model.

2.3.1. Phenology-based wheat mapping

Although the phenological difference is relatively large across different regions, the morphological characteristics of phenological curves exhibit similarity. More specifically, winter wheat is generally

sown in autumn, undergoes a dormancy in winter, followed by a green-up phase in early spring of the sequent year (only for the Northern Hemisphere), and then grows rapidly up until the heading date (late spring) and finally reaches maturity in early summer. In contrast, spring wheat has a much shorter growth period that spans from spring to late summer. Detecting the distinctive features of phenological stages in LAI temporal profiles contributes to distinguish different crops. For example, winter and spring wheat can be easily distinguished according to the fact that the heading (maximum point of the LAI time series) and senescence phase of winter wheat occur earlier than those of spring wheat (Fig. S16). Moreover, heading and senescence of wheat frequently occur at least 30 days earlier than other summer crops (e.g., maize and soybean). In the study, we modified the method used by Luo et al. (2020) and developed a wheat detecting algorithm to automatically detect three “wheat signals” as follows:

- 1) The inflection point (green-up/emergence) is determined by the occurrence of at least three positive first derivatives in a temporal window of 40-day composites after this point;
- 2) The maximum point necessitates satisfying the three rules below: occurs in a temporal-specific window (Win_{max}) that is derived from previous studies on wheat phenology; it is preceded (followed) by a continuous increase (decrease) in LAI that is identified as at least three positive (negative) first derivatives in a temporal window of 40-day composites; and the maximum LAI is above a specified threshold of LAI_{max} .

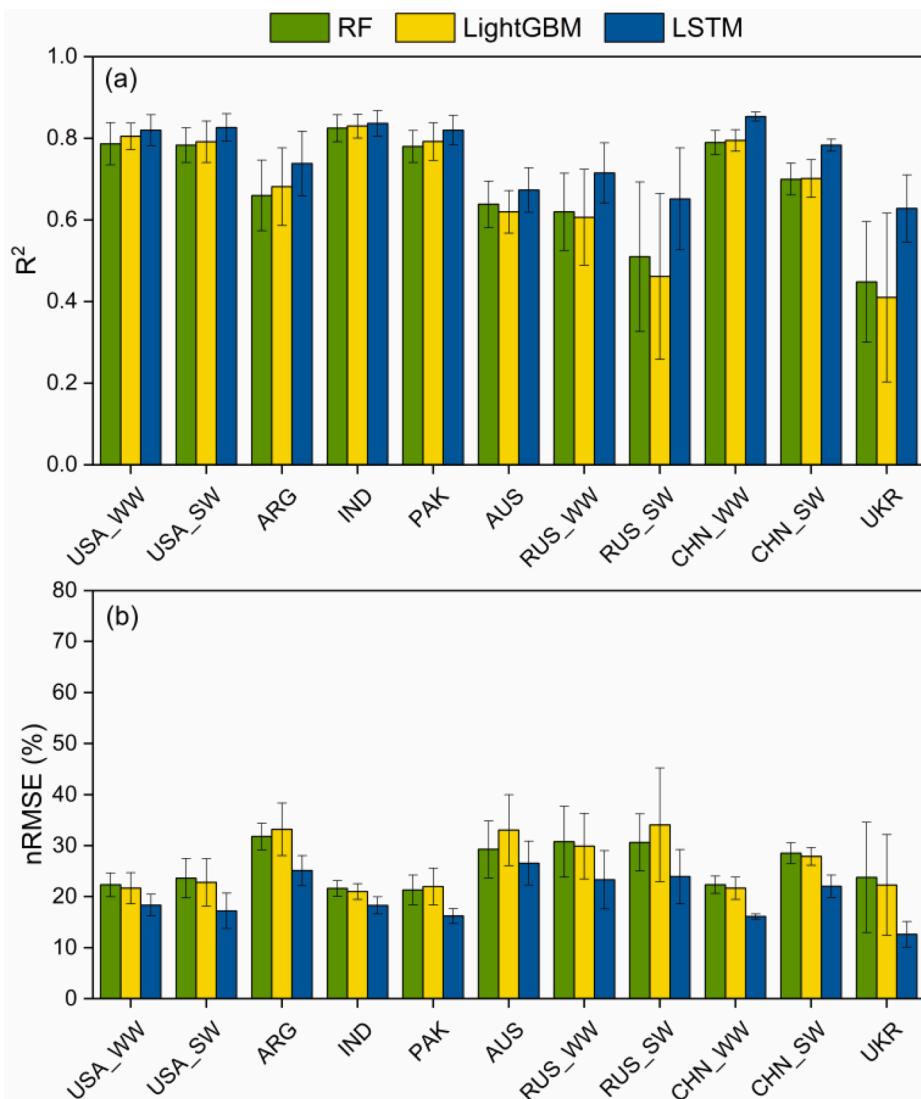


Fig. 4. Performance of three different methods in yield estimation during 2007–2012: (a) R², (b) nRMSE. USA_WW: winter wheat in USA; USA_SW: spring wheat in USA; RUS_WW: winter wheat in RUS; RUS_SW: spring wheat in RUS; CHN_WW: winter wheat in China; CHN_SW: spring wheat in China.

3) The senescence phase (an abrupt decrease in LAI) is detected as a drop of at least LAI_{dec}% in the LAI within a temporal window of 40-day composites.

We first used the cropland layer derived from the GFSAD1km as the cropland mask. Then, we applied the prevalent Savitzky-Golay (S-G) filter method to reconstruct GLASS LAI time series, which has been suggested to be a trustworthy tool (Luo et al., 2020; Savitzky and Golay, 1964). Finally, we implemented the algorithm to generate annual harvesting area of spring and winter wheat for the years 2007–2012. More specifically, the pixels that simultaneously meet the above-mentioned rules are labeled as those with wheat cultivated.

2.3.2. Developing the optimal grid-scale yield estimation model

To investigate the best yield estimation model for generating accurate gridded yield maps, we compared three prevalent DL and ML algorithms (i.e., RF, LightGBM, and LSTM). Details for the three models were given in [Supplementary Text 1](#). To reduce dimensions and select the most appropriate variables for yield estimation across different regions, exploratory data analysis was conducted carefully before developing the three models (see [Supplementary Text 2](#)). First, we resampled the NDVI, climatic, and soil data into 4 km using nearest neighbor

method. Moreover, monthly NDVI composites were derived from the maximum value composite method. Next, we determined the annual wheat cultivated areas on the condition that the grids were cultivated for at least one/two year during adjacent two/three years to reduce inevitable uncertainties of remote sensing data. The integrated wheat maps were also resampled into 4 km. Ultimately, we aggregated the inputs at the subnational scale by calculating the mean after being masked by the final wheat map. The above-mentioned procedures were implemented by the Google Earth Engine (GEE) platform.

2.3.3. Model evaluation

We implemented “leave-one-year-out” strategy to eliminate temporal autocorrelation among samples. More specifically, we recursively chose each year from 2007 to 2012 for testing, and the remaining years’ data for training. For example, the time coverage of recorded yield in China at a county level was from 1982 to 2015. Taking 2010 for testing as an example, three models were first trained with the data for the years 1982–2009 and 2011–2015. The hyper-parameters for each model were determined based on the five-fold cross-validated coefficient of determination (R²) using GridSearchCV packages with the training data alone (Cai et al., 2019). Next, we implemented the three optimized models to derive gridded yield estimates in 2010. Finally, we averaged the gridded

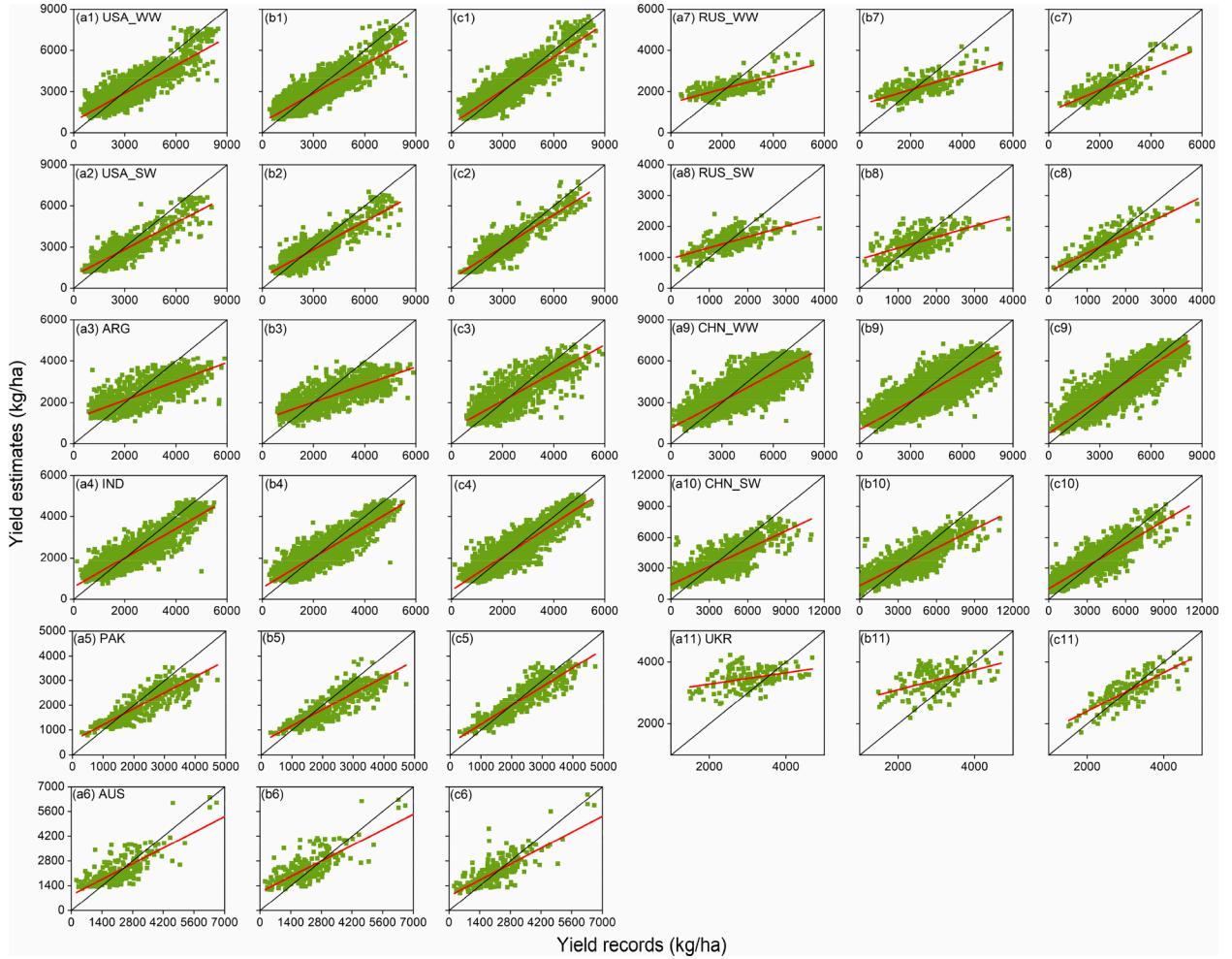


Fig. 5. Comparisons between the observed and estimated yield by (a) RF, (b) LightGBM, (c) LSTM during 2007–2012. USA_WW: winter wheat in USA; USA_SW: spring wheat in USA; RUS_WW: winter wheat in RUS; RUS_SW: spring wheat in RUS; CHN_WW: winter wheat in China; CHN_SW: spring wheat in China.

yields by county and then calculated the root mean square error (RMSE) and R^2 using the subnational statistics for 2010. The whole procedure was repeated 10 times to dampen uncertainty and evaluate the practical prediction performance of three methods.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

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

$$\text{nRMSE} = \frac{\text{RMSE}}{\bar{y}} \quad (3)$$

where y_i and \hat{y}_i are the recorded and estimated yield, individually. \bar{y} implies the average of recorded yield.

2.3.4. Investigating the advantages of GWPMS

SPAM utilized a cross-entropy approach to estimate crop harvesting area and yield at a spatial resolution of 5-arcmin for the years 2000, 2005, and 2010, by integrating multi-source data such as crop production statistics, land cover maps, irrigated areas and cropping intensity (You et al., 2014). First, we compared the consistency of yield estimates derived from SPAM or GWPMS with statistics for 2000, 2005, and 2010. To further substantiate the advantages of GWPMS in accurately mapping finer-resolution yields at the global scale, we calculated the R^2 and

RMSE between the statistical and predicted yields based on the wheat map derived from SPAM and GWPMS from 2007 to 2012, separately.

3. Results

3.1. Accuracy of wheat-area maps

Overall, the estimated wheat harvesting areas agreed well with the statistics as they were evenly distributed along the 1:1 line with R^2 (0.78 ~ 0.87) and RMSE (8.1 ~ 235.2 Kha) during 2007–2012 (Fig. 3, Table S2), except for spring wheat in the USA, AUS and RUS with harvesting areas overestimated. We attributed such deviations of spring wheat to interferences from other spring cereals (e.g., spring barley and oats); that is, they have similar spectral signatures. In contrast, such interferences were relatively weaker for winter wheat during the detection processes. Furthermore, we compared the resultant wheat maps with the CDL products (Fig. S2). The spatial distribution of the results matched with that of the CDL, with R^2 (RMSE) equal to 0.82 (6.2 Kha) and 0.81 (21.4 Kha) for winter and spring wheat, respectively. Therefore, our derived maps were reliable because of their high consistencies with the agricultural census data and popular crop map for both harvesting areas and spatial distribution.

3.2. Comparisons of three models for yield estimates at the grid scale

The prediction performance of three ML and DL models during

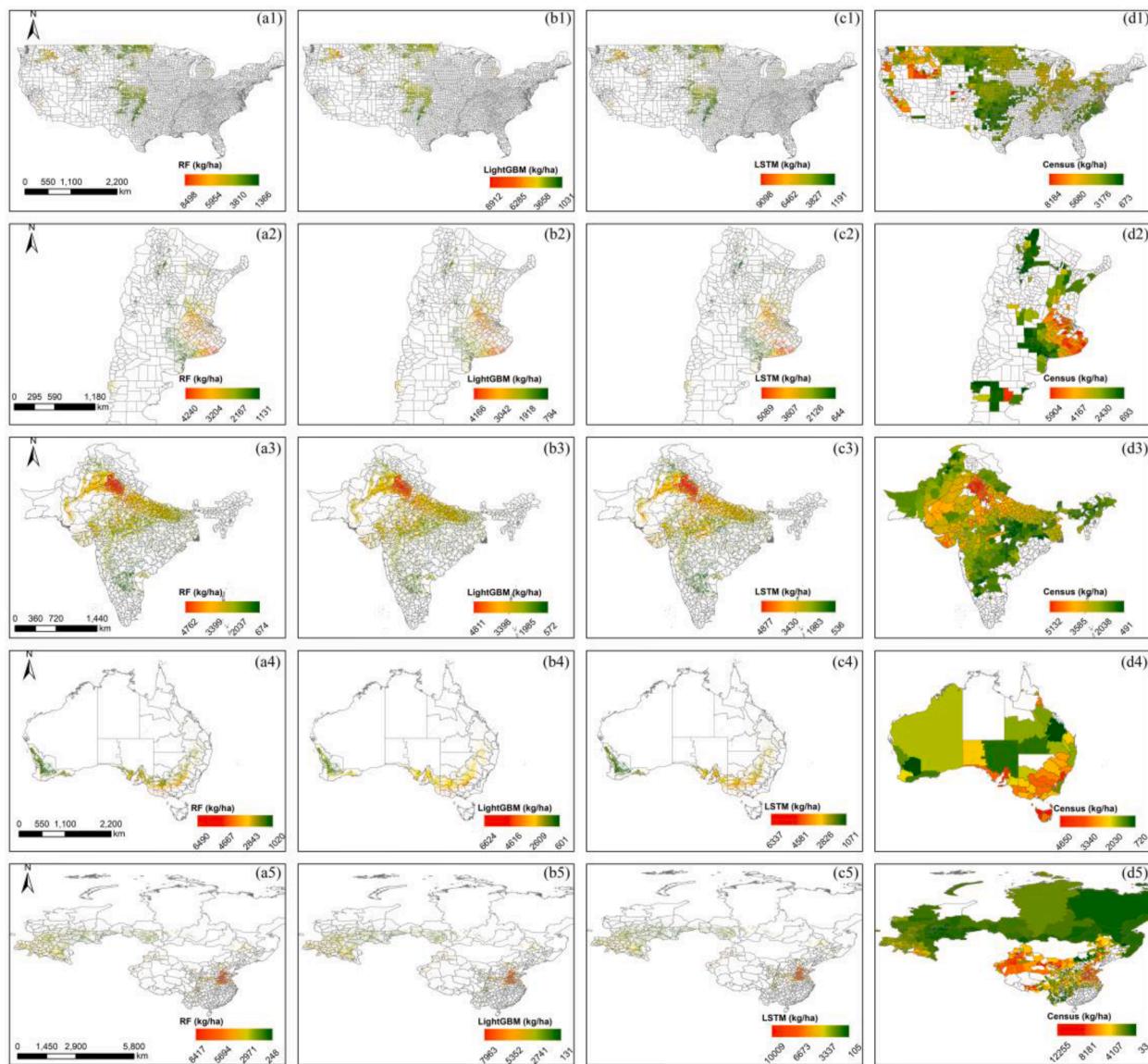


Fig. 6. Spatial patterns of the estimated yield for 2010 by RF (a), LightGBM (b), and LSTM (c), as well as the subnational-level yield statistics in 2010 (d).

2007–2012 for each country at the grid scale were shown in Figs. 4–5 and Table S3. Overall, LSTM generally outperformed RF and LightGBM with average R^2 (nRMSE) of 0.76 (19.8), 0.7 (25.9), and 0.69 (26.3), respectively (Fig. 4, Table S3). The LSTM model performed the best in China, the United States, India and Pakistan, with R^2 consistently >0.8 . In this case, both the ML methods achieved comparable accuracies (R^2 between 0.7 and 0.83). In addition, LSTM provided a substantial improvement in R^2 by $\sim 13.5\%$, especially in Argentina, Australia, Russian Federation and Ukraine with improvements ranging from 10% to 46%. Fig. 5 showed that the predicted yields of LSTM averaged at the subnational scale were more consistent with recorded yields, indicated by the fitted lines close to the 1:1 line than those of two ML methods. The better performance of LSTM further suggests LSTM can avoid excessive underestimation (overestimation) of high- (low-) yield cases as two ML methods do (Fig. 5). Such superiority is typically reflected in some countries such as Argentina, Ukraine and Russian Federation, possibly ascribed to its stronger temporal learning ability that can recognize complex sequential patterns over long time periods. In addition, the spatial patterns of the estimated yield for 2010 by three methods were generally consistent with that of the statistical yields, particularly for the LSTM models (Fig. 6). The comparison also demonstrated that the LSTM

models could better capture the high-yield cases than ML methods do.

Fig. S14 exhibited the distribution of estimation errors for each country in 2010. Totally, the LSTM models reduced bias compared to the ML models, particularly in Argentina, India, Pakistan, the United States, Russian Federation and Ukraine. In contrast, the estimation errors of both ML models were mainly distributed on the negative side and were greater than those of LSTM, suggesting more underestimations and worse performances for subnational-level yields.

The LSTM model was chosen as the best model for estimating yield in GWPMS because of its better performance than the two ML models. We further evaluated the temporal consistency between the time series of estimated yields by GWPMS and those observed (Fig. S15). The average correlation coefficients ranged from 0.4 to 0.7, implying a strong temporal learning ability of GWPMS.

3.3. Comparing the performance of GWPMS and SPAM in mapping wheat production

We compared the consistencies of yield maps generated by GWPMS and SPAM in 2000, 2005 and 2010 with the census yields, respectively. Generally, the yield maps derived from GWPMS showed a higher

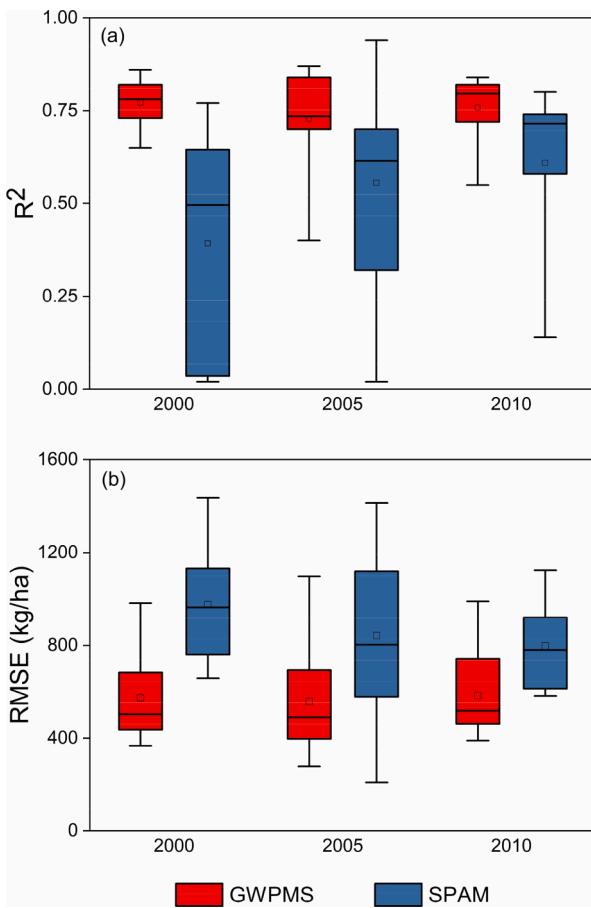


Fig. 7. Boxplot of subnational-level comparisons between the census yield and estimated yield by GWPMS or SPAM in 2000, 2005, and 2010: (a) R^2 , (b) RMSE.

accuracy than SPAM among all the years and countries with averaged R^2 (RMSE) of 0.75 (572.2 Kha) and 0.52 (871.5), respectively (Fig. 7, Table 1). In addition, the accuracies of GWPMS-derived yield maps were more stable among the three years while those of SPAM datasets varied largely with R^2 ranging from 0.39 to 0.61, potentially due to the continuous improvements of SPAM in methodology and input data. Compared with SPAM, the increases (decreases) in R^2 (RMSE) for GWPMS were approximately 30% and varied by years and countries. The largest improvements were found in Ukraine, Australia and Pakistan with R^2 (RMSE) increased (decreased) by > 150% (>27%), perhaps owing to the more accurate spatial distribution of wheat and consistently high-resolution input data.

Furthermore, we compared the performances of LSTM models based on different wheat maps derived from GWPMS and SPAM from 2007 to 2012, respectively (Fig. 8). Overall, the R^2 (RMSE) was improved (reduced) by an average of 6.7% (6.3%) when using GWPMS-derived maps instead of SPAM, especially in Argentina, China, Ukraine, Russian Federation and the United States. Such comparisons further demonstrate that more accurate spatial information on crops will contribute to a better performance for yield estimation at a grid scale. Unfortunately, the issue has been ignored by most peers in the field. Alternatively, an unchanged crop map has been popularly used for predicting crop yields over a long-term period.

4. Discussion

4.1. Accurately dynamic crop distribution map improves yield estimation

We developed a generalized GWPMS approach to map wheat production based on the LSTM models, using the subnational statistics with more political units and a longer time period than any previous study, stepwise from harvesting area to yield. Few previous works thus far have paid particular attention to mapping global crop yields using data-driven models based on accurate crop distributions, although some of them devoted to the county- or field-level yield estimation (Cao et al., 2021; Schwalbert et al., 2020; Zhang et al., 2021). One of the strengths of GWPMS was that it offered a trustworthy tool for accurately mapping spatial distribution of wheat, which could be replicated over large spatial extents without the need for training samples of different land covers that are difficult to collect continuously. The accurate and annual crop maps were essential to estimate yield due to their key roles in ensuring the robustness of aggregated features over the statistical units that reflect crop growth status (Cai et al., 2019). Our study has proved that the accurate and dynamic wheat maps have improved yield estimates by an R^2 of 6.7% compared with the coarse maps of SPAM (Fig. 8). The finding further strongly substantiated higher-resolution crop map will benefit yield estimation at the grid scale, implying a potential yield over/underestimation from most previous works that have utilized a constant crop map (Cai et al., 2019; Iizumi et al., 2014b).

4.2. Deep learning algorithm outperform machine learning in estimating crop yields

Our GWPMS demonstrated that DL such as the LSTM model provides a powerful data-driven approach to estimate gridded wheat yield at the global scale. LSTM consistently outperformed two ML models (i.e., RF and LightGBM) regardless of year and country, especially enabling better capturing higher-yield cases (Figs. 5 and 6). The primary reasons for the superiority of LSTM were that its recurrent neural network could characterize the complex and nonlinear relationships between high-

Table 1
Comparisons of yield between the GWPMS- or SPAM-based estimates and subnational-level statistics.

Country/Crop type	2000				2005				2010			
	GWPMS		SPAM		GWPMS		SPAM		GWPMS		SPAM	
	R^2	RMSE (kg/ha)										
USA_WW	0.86	555.8	0.77	657.9	0.77	693.2	0.55	1075.9	0.8	579.1	0.58	871.9
USA_SW	0.77	645.9	0.74	1202.5	0.7	725.4	0.63	1340.1	0.82	512.1	0.74	807.9
ARG	0.65	424.5	0.02	875.6	0.74	497.3	0.66	578.5	0.72	748.7	0.71	751.2
IND	0.83	448.2	0.49	841.8	0.87	396.2	0.7	613.2	0.83	461.7	0.8	588.1
PAK	0.77	447.7	0.04	1052.4	0.84	278.3	0.32	885.1	0.81	388.9	0.63	737.3
AUS					0.7	480.2	0.29	1118.1	0.65	522	0.14	920.7
RUS	0.69	366	0.03	680.4	0.73	416.5	0.94	209.1	0.79	460.5	0.72	613
CHN_WW	0.79	718.6	0.55	1059.3	0.84	684.1	0.85	720	0.84	742.9	0.8	975.2
CHN_SW	0.81	983	0.5	1435	0.7	1097.6	0.6	1412.5	0.77	990.2	0.73	1123.5
UKR					0.4	316	0.02	464.4	0.55	437.9	0.25	582.5
Mean	0.77	573.7	0.39	975.6	0.73	558.5	0.56	841.7	0.76	584.4	0.61	797.1

USA_WW: winter wheat in USA; USA_SW: spring wheat in USA; CHN_WW: winter wheat in China; CHN_SW: spring wheat in China.

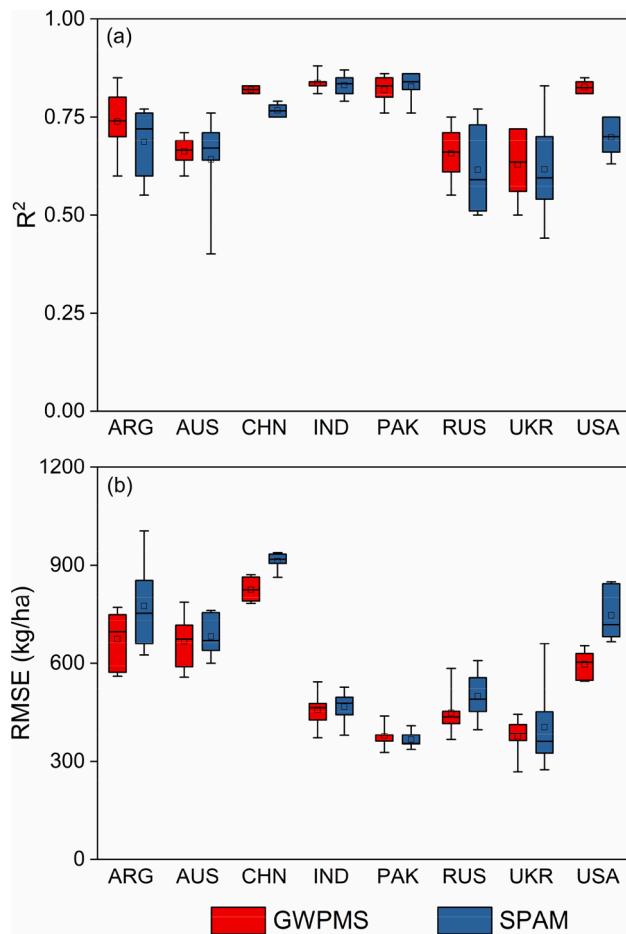


Fig. 8. Boxplot of subnational-level comparisons between the predicted yield by the LSTM model and census yield from 2007 to 2012 based on wheat map derived from GWPMS or SPAM: (a) R^2 , (b) RMSE.

dimensional (spectral, temporal, and spatial) geospatial data and yield, and leverage temporal cumulative impact of climatic factors on crop yield (Jiang et al., 2020). Consequently, the wheat yield maps generated by GWPMS showed a higher and more robust consistency with agricultural census data in comparisons with SPAM over space and time (Fig. 7). In general, GWPMS provides a scalable, automatic and systematic framework for accurately mapping global wheat production information using publicly available data.

4.3. Some limitations of GWPMS

Although GWPMS achieved a highly credible accuracy in mapping wheat production information, there are still some limitations. First, we used a single wheat map that masks gridded input data by integrating annual maps during 2007–2012 to estimate wheat yields for the years before 2007 because GFSAD1km was merely produced during this period. For the six annual maps 2007–2012, the integrating maps of adjacent two/three years were used for masking. Thus our wheat maps have not really characterized the dynamic variability in wheat-planting areas over time partly from avoiding the uncertainties of remote sensing data. The uncertainties were mainly caused by the coarse spatial resolution of 1 km, which could potentially weaken the temporal signature of winter and spring wheat, especially for areas with many mixed pixels and low crop density. For example, these pixels can dampen the “senescence signal” (e.g., delay the senescence phase), consequently resulting in a lower decrease in LAI than the determined threshold and ultimately failed in being identified as wheat. Till nowadays, however, consecutively monitoring of crop distribution over a large spatial extent

is still challenging due to data availability. With more remote sensing images at finer spatial resolutions available, we are sure more accurate crop maps will be provided (e.g., Harmonized Landsat Sentinel-2 data) to avoid mixed pixel issues, as well as other promising methods for large-scale applications such as transfer learning in the near future (Kluger et al., 2021).

Additionally, the LSTM model had the disadvantages of its ‘black box’ property, suggesting poor traceability and interpretability (Jiang et al., 2020). Therefore, more state-of-the-art analytical tools such as attention (AT) based network (e.g., AT-LSTM) is expected to interpret DL-based crop yield estimation (Lin et al., 2020). Moreover, many novel networks such as Spatial Variability Aware Deep Neural Networks (<https://arxiv.org/abs/2011.08992>) have shown trustworthy performance in capturing spatial-temporal variability, and their ability to map global crop yield should be further investigated. Ultimately, the spatiotemporal scalability of the LSTM models should be investigated in the future because of the limited or even scarce statistical data in some regions (e.g., Ukraine, Middle East, and Africa), through applying the models trained in regions or years with plentiful data into a target region or year where census data are lacking (You et al., 2017).

5. Conclusion

We demonstrated a generalized approach (GWPMS) to map global wheat production stepwise from harvesting area to yield. First, we integrated the inflexion- and threshold-based methods to generate wheat harvesting area maps for eight major wheat-producing countries from 2007 to 2012. Then, we used the accurate maps to derive aggregated indices over the subnational scale on the GEE platform (including satellite, climatic data and soil properties) and compared the performance of three ML/DL algorithms for estimating gridded wheat yield. Finally, we compared our resultant area and yield maps with those of popular products (e.g. SPAM) to demonstrate the superiority of GWPMS. Overall, the classification results showed a high accuracy as compared with the statistics ($R^2 > 0.78$). In addition, our wheat maps were in strong consistency with the CDL ($R^2 > 0.8$). The LSTM model achieved the highest accuracy than RF and LightGBM in gridded yield estimates, with R^2 (nRMSE) ~ 0.76 (19.8), 0.7 (25.9), and 0.69 (26.3), respectively. Compared with SPAM ($R^2 \sim 0.52$), the final yield maps had a higher and more robust accuracy with an average R^2 of 0.75 across all countries and years. The more accurate wheat maps derived from GWPMS improved the predictive performance of the LSTM model with R^2 (RMSE) increased (decreased) by an average of 6.7% (6.3%) in comparisons with those from SPAM. Our findings demonstrate a scalable, automatic and trustworthy framework for mapping wheat production with publicly available data, which is promising to be easily applied over the globe. This approach and generated datasets have important implications for many applications, such as large-scale agricultural systems modeling, crop insurance designs and policy-making.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by National Natural Science Foundation of China (42061144003, 41977405).

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jag.2022.102823>.

References

- Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* 5, 170191. <https://doi.org/10.1038/sdata.2017.191>.
- Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sens. Environ.* 202, 129–141. <https://doi.org/10.1016/j.rse.2017.04.014>.
- Cai, Y.P., Guan, K.Y., Lobell, D., Potgieter, A.B., Wang, S.W., Peng, J., Xu, T.F., Asseng, S., Zhang, Y.G., You, L.Z., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. Forest Meteorol.* 274, 144–159. <https://doi.org/10.1016/j.agrformet.2019.03.010>.
- Cao, J., Zhang, Z., Luo, Y.C., Zhang, L.L., Zhang, J., Li, Z.Y., Tao, F.L., 2021. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *Eur. J. Agron.* 123, 126204 <https://doi.org/10.1016/j.eja.2020.126204>.
- de Wit, A.J.W., van Diepen, C.A., 2008. Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. *Int. J. Appl. Earth Obs.* 10 (4), 414–425. <https://doi.org/10.1016/j.jag.2007.10.004>.
- Deines, J.M., Patel, R., Liang, S.Z., Dado, W., Lobell, D.B., 2021. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn. *Remote Sens. Environ.* 253, 112174 <https://doi.org/10.1016/j.rse.2020.112174>.
- FAO/IIASA/ISRIC/ISSCAS/JRC, 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.
- FAO, 2019. Available at: <http://www.fao.org/faostat/zh/#data/CC>.
- Folberth, C., Skalsky, R., Moltchanova, E., Balkovic, J., Azevedo, L.B., Obersteiner, M., van der Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nat. Commun.* 7, 11872. <https://doi.org/10.1038/ncomms11872>.
- Gomez, D., Salvador, P., Sanz, J., Casanova, J.L., 2021. Modelling wheat yield with antecedent information, satellite and climate data using machine learning methods in Mexico. *Agric. Forest Meteorol.* 300, 108317 <https://doi.org/10.1016/j.agrformet.2020.108317>.
- Iizumi, T., Luo, J.J., Challinor, A.J., Sakurai, G., Yokozawa, M., Sakuma, H., Brown, M.E., Yamagata, T., 2014a. Impacts of El Nino Southern Oscillation on the global yields of major crops. *Nat. Commun.* 5, 3712. <https://doi.org/10.1038/ncomms4712>.
- Iizumi, T., Yokozawa, M., Sakurai, G., Travassos, M.I., Romanenkov, V., Oettli, P., Newby, T., Ishigooka, Y., Furuya, J., 2014b. Historical changes in global yields: major cereal and legume crops from 1982 to 2006. *Glob. Ecol. Biogeogr.* 23 (3), 346–357. <https://doi.org/10.1111/geb.12120>.
- Jeong, S., Ko, J., Yeom, J.M., 2022. Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. *Sci. Total Environ.* 802, 149726 <https://doi.org/10.1016/j.scitotenv.2021.149726>.
- Jiang, H., Hu, H., Zhong, R.H., Xu, J.F., Xu, J.L., Huang, J.F., Wang, S.W., Ying, Y.B., Lin, T., 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. Change Biol.* 26 (3), 1754–1766. <https://doi.org/10.1111/gcb.14885>.
- Jin, Z.N., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* 228, 115–128. <https://doi.org/10.1016/j.rse.2019.04.016>.
- Kluger, D.M., Wang, S., Lobell, D.B., 2021. Two shifts for crop mapping: Leveraging aggregate crop statistics to improve satellite-based maps in new regions. *Remote Sens. Environ.* 262, 112488 <https://doi.org/10.1016/j.rse.2021.112488>.
- Lai, Y.R., Pringle, M.J., Kopitke, P.M., Menzies, N.W., Orton, T.G., Dang, Y.P., 2018. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *Int. J. Appl. Earth Obs.* 72, 99–108. <https://doi.org/10.1016/j.jag.2018.07.013>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444. <https://doi.org/10.1038/nature14539>.
- Li, Y., Guan, K.Y., Yu, A., Peng, B., Zhao, L., Li, B., Peng, J., 2019. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S. *Field Crop Res.* 234, 55–65. <https://doi.org/10.1016/j.fcr.2019.02.005>.
- Lin, T., Zhong, R.H., Wang, Y.D., Xu, J.F., Jiang, H., Xu, J.L., Ying, Y.B., Rodriguez, L., Ting, K.C., Li, H.F., 2020. DeepCropNet: a deep spatial-temporal learning framework for county-level corn yield estimation. *Environ. Res. Lett.* 15 (3), 034016 <https://doi.org/10.1088/1748-9326/ab66cb>.
- Liú, H., Gong, P., Wang, J., Clinton, N., Bai, Y.Q., Liang, S.L., 2020. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth Syst. Sci. Data* 12 (2), 1217–1243. <https://doi.org/10.5194/essd-12-1217-2020>.
- Lobell, D.B., Burke, M., 2010. On the use of statistical models to predict crop yield responses to climate change. *Agric. Forest Meteorol.* 150 (11), 1443–1452. <https://doi.org/10.1016/j.agrformet.2010.07.008>.
- Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* 164, 324–333. <https://doi.org/10.1016/j.rse.2015.04.021>.
- Luo, Y.C., Zhang, Z., Li, Z.Y., Chen, Y., Zhang, L.L., Cao, J., Tao, F.L., 2020. Identifying the spatiotemporal changes of annual harvesting areas for three staple crops in China by integrating multi-data sources. *Environ. Res. Lett.* 15 (7), 074003 <https://doi.org/10.1088/1748-9326/ab80f0>.
- Luo, Y.C., Zhang, Z., Zhang, L.L., Cao, J., 2021. Spatiotemporal patterns of winter wheat phenology and its climatic drivers based on an improved pDSSAT model. *Sci. China Earth Sci.* 64 (12), 2144–2160. <https://doi.org/10.1007/s11430-020-9821-0>.
- Nelson, G.C., Rosegrant, M.W., Palazzo, A., Gray, I., Ingersoll, C., Robertson, R., et al., 2010. Food security, farming, and climate change to 2050: Scenarios, results, policy options. IFPRI, Washington.
- Nelson, K.S., Burchfield, E.K., 2021. Landscape complexity and US crop production. *Nat. Food* 2 (5), 330–338. <https://doi.org/10.1038/s43016-021-00281-1>.
- Peng, B., Guan, K.Y., Zhou, W., Jiang, C.Y., Frankenberg, C., Sun, Y., He, L.Y., Kohler, P., 2020. Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction. *Int. J. Appl. Earth Obs.* 90, 102126 <https://doi.org/10.1016/j.jag.2020.102126>.
- Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A., 2012. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3, 1293. <https://doi.org/10.1038/ncomms2296>.
- Rembold, F., Atzberger, C., Savin, I., Rojas, O., 2013. Using Low Resolution Satellite Imagery for Yield Prediction and Yield Anomaly Detection. *Remote Sens.-Basel* 5 (4), 1704–1733. <https://doi.org/10.3390/rs5041704>.
- Sakamoto, T., 2020. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS J. Photogramm. Remote Sens.* 160, 208–228. <https://doi.org/10.1016/j.isprsjprs.2019.12.012>.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing C Differentiation Of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V., Ciampitti, I.A., 2020. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. Forest Meteorol.* 284, 107886 <https://doi.org/10.1016/j.agrformet.2019.107886>.
- Xiao, Z.Q., Liang, S.L., Wang, J.D., Xiang, Y., Zhao, X., Song, J.L., 2016. Long-Time-Series Global Land Surface Satellite Leaf Area Index Product Derived From MODIS and AVHRR Surface Reflectance. *IEEE Trans. Geosci. Remote* 54 (9), 5301–5318. <https://doi.org/10.1109/TGRS.2016.2560522>.
- Yadav, K., Congalton, R.G., 2018. Accuracy Assessment of Global Food Security-Support Analysis Data (GFSAD) Cropland Extent Maps Produced at Three Different Spatial Resolutions. *Remote Sens.-Basel* 10 (11), 1800. <https://doi.org/10.3390/rs10111800>.
- You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep gaussian process for crop yield prediction based on remote sensing data. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1.
- You, L.Z., Wood, S., Wood-Sichra, U., Wu, W.B., 2014. Generating global crop distribution maps: From census to grid. *Agric. Syst.* 127, 53–60. <https://doi.org/10.1016/j.agsy.2014.01.002>.
- Zhang, L.L., Zhang, Z., Luo, Y.C., Cao, J., Xie, R.Z., Li, S.K., 2021. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. *Agric. Forest Meteorol.* 311, 108666 <https://doi.org/10.1016/j.agrformet.2021.108666>.