



# Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine



Juan Cao<sup>a</sup>, Zhao Zhang<sup>a,\*</sup>, Yuchuan Luo<sup>a</sup>, Liangliang Zhang<sup>a</sup>, Jing Zhang<sup>a</sup>, Ziyue Li<sup>a</sup>, Fulu Tao<sup>b,c</sup>

<sup>a</sup> State Key Laboratory of Earth Surface Processes and Resources Ecology, Key Laboratory of Environmental Change and Natural Disaster MOE, Faculty of Geographical Science, Beijing Normal University, Beijing, China

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

<sup>c</sup> College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, China

## ARTICLE INFO

### Keywords:

Winter wheat  
Machine learning  
Deep learning  
Google earth engine (GEE)  
Yield estimation

## ABSTRACT

To meet the challenges of climate change, increasing population and food demand, a timely, accurate and reliable estimation of crop yield at a large scale is more imperative than ever for crop management, food security evaluation, food trade and policy-making. In this study, taking the major winter wheat production regions of China as an example, we compared a traditional machine learning method (random forest, RF) and three deep learning (DL) models, including DNN (deep neural networks), 1D-CNN (1D convolutional neural networks), and LSTM (long short-term memory networks) to predict crop yields by integrating publicly available data within the GEE (Google Earth Engine) platform, including climate, satellite, soil properties, and spatial information data. The results showed that all four models could capture winter wheat yield variations in all the county-years, with  $R^2$  of recorded and simulated yields ranging from 0.83 to 0.90 and RMSE ranging from 561.18 to 959.62 kg/ha. They all performed well for winter wheat yield prediction at a county level from 2011 to 2015, with mean  $R^2 \geq 0.85$  and  $RMSE \leq 768$  kg/ha. At a field level, the spatial pattern of estimated winter wheat yield could capture the spatial heterogeneity and yield differences between individual fields across a county fairly well. However, only the DNN and RF models had relatively good performance at the field level, with mean  $R^2$  values of 0.71, 0.66 and RMSE values of 1127 kg/ha and 956 kg/ha, respectively. The model comparisons showed that the performance of RF was not always worse than DL at both the county and field levels. Our findings demonstrated a scalable, simple and inexpensive framework for estimating crop yields at various scales in a timely manner and with reliable accuracy, which has important implications for crop yield forecasting, agricultural disaster monitoring, food trade policy, and food security warning.

## 1. Introduction

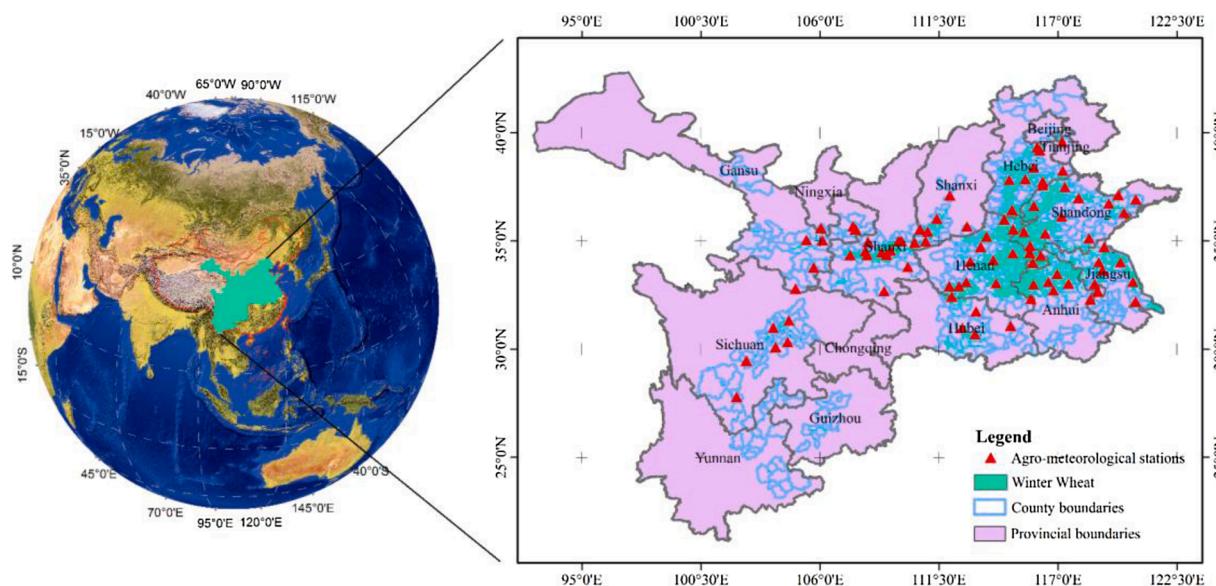
Approximately 795 million people worldwide still live without an adequate food supply (FAO, 2015), and by 2050, there will be two billion more people to feed and an approximately double global demand for food (Dodds and Bartram, 2016; Gorelick et al., 2017; You et al., 2017). Approximately 35–40 % of the world's population lives on wheat, but the production of wheat has stagnated in approximately 37 % of the wheat cultivation areas worldwide in recent years (Ray et al., 2012). China is the world's largest producer and consumer of wheat (He et al.,

2010) and will face more challenges in meeting increasing food demands. Winter wheat accounts for approximately 85 % of China's total summer grain production (Huang et al., 2015). However, wheat yields stagnated in 56 % of China's areas from 1961 to 2008 (Chen et al., 2017; Ray et al., 2012). Therefore, a timely, accurate and spatial prediction of wheat yields is of great importance to ensure farmers' interests and national food security in China.

Satellite remote sensing provides strong advantages over other monitoring techniques by providing a timely, synoptic, and up-to-date overview of large-scale crop monitoring at multiple stages (Guan

\* Corresponding author at: State Key Laboratory of Earth Surface Processes and Resource Ecology, Key Laboratory of Environmental Change and Natural Hazards MOE, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China.

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

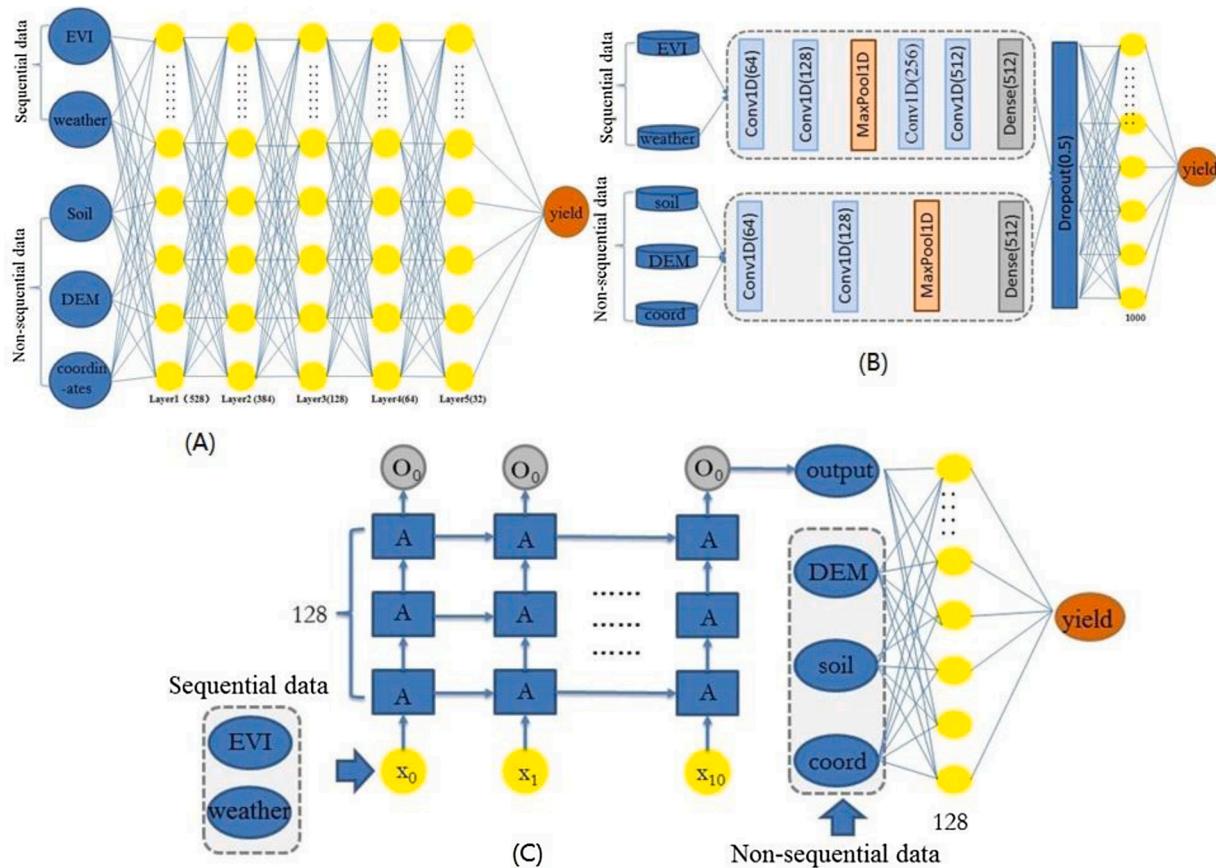


**Fig. 1.** The locations of study area, counties and the agro-meteorological stations used in this study.

et al., 2017; Yang et al., 2019). Satellite-based, remote sensing crop monitoring techniques (which can acquire spatiotemporal data across large spatial scales) have been successfully employed for crop yield estimation in the past few decades (Lobell et al., 2015; Macdonald and Hall, 1980; Sakamoto et al., 2013; Yang et al., 2019). The first crop yield estimation using remote sensing data was proposed in the late 1970s. Macdonald and Hall (1980) estimated harvests in counties of strategic interest. The existing remote sensing yield estimation methods have been mainly based on empirical relationships between vegetation indices (VIs) and field-measured yields. The main drawback to these methods is that the models are only applicable to specific crop cultivars, locations and years and not to a large-scale geographical region, which could create problems with the extrapolation of equations to another year and location. Therefore, new field-measurement yields are also needed for each new setting (Huang et al., 2015; Lobell, 2013; Lobell et al., 2015; Sakamoto et al., 2013; Yang et al., 2019). In addition, process-oriented crop simulation models (e.g., DSSAT, APSIM, WOFOST, MCWLA, and AQUACROP) or crop simulation models assimilated with remote sensing data can better accommodate changes in location, weather and timing of images and estimate crop yield for each pixel (Battisti et al., 2017; Huang et al., 2015; Lobell et al., 2015), but they often need more detailed input data, such as site-specific soil and daily weather data. Both the computational costs and data requirements can hinder scaling the approach to multiple crops, regions and years without a significant investment of time, money, and labour (Lobell, 2013; Lobell et al., 2015).

Recently, the cost of obtaining and pre-processing imagery has been drastically lowered due to the widespread availability of imagery and state-of-the-art cloud-based platforms for planetary-scale geospatial analysis (e.g., Google Earth Engine, GEE) (Gorelick et al., 2017). The GEE platform provides a centralized warehouse of petabytes of earth observation data and a library of functions for executing user-supplied analyses. The data warehouse archives a large catalogue of satellite imagery, meteorological data, elevation data and atmospheric data. Users can also upload their data for analysis. Analyses run on these data are automatically parallelized to run across many CPUs in the Google cloud, dramatically decreasing the cost of time and labour to complete the computations. These massive computational capabilities have been used to research a variety of high-impact societal issues, such as global forest change (Hansen et al., 2013), surface water change and mapping (Pekel et al., 2016), crop yield mapping (Azzari et al., 2017; Lobell et al., 2015), flood mapping (Coltin et al., 2016), rice paddy extraction (Dong

et al., 2016; Singha et al., 2019), disease risk mapping (Sturrock et al., 2014), climate monitoring (Gorelick et al., 2017) tracking changes in tiger habitats (Joshi et al., 2016), and crop classification (Cai et al., 2018). Furthermore, recent rapid development in artificial intelligence has led to an increase machine learning (ML) and deep learning (DL) algorithms, which have been successfully applied in various domains and obviously outperform other traditional techniques (LeCun et al., 2015). Therefore, some recent studies have applied DL and ML for yield estimation (Johnson, 2014; Khaki and Wang, 2019; Kuwata and Shibasaki, 2015; Russello, 2018; Sakamoto et al., 2013; Yang et al., 2019; You et al., 2017), including support vector machines (SVMs), deep neural networks (DNNs), convolutional neural networks (CNNs), and long-short-term memory (LSTM). For example, Kuwata and Shibasaki (2015) applied SVM and a two inner-product layer neural network to estimate maize yields. You et al. (2017) used CNN and LSTM to estimate soybean yields in the United States. The results outperformed traditional remote-sensing-based methods and USDA national-level estimation. Russello (2018) employed 3-dimensional CNN for soybean yield prediction based on satellite images. The results were better than other traditional techniques. Khaki and Wang (2019) designed a DNN to predict the maize yield performance in 2017 based on field-measurement data from 2008 to 2016, which showed that the model had a superior prediction accuracy than Lasso, shallow neural networks (SNNs), and regression trees (RTs). Yang et al. (2019) trained a CNN method by very high-resolution UAV-based remote sensing images to predict field rice yield. They claimed that the results were much better than the traditional VI-based regression model. Nevertheless, the application of ML and DL to crop yield estimation is in its infant stage. In particular, the potential of application of publicly available datasets and different ML and DL methods for crop yield estimations across scales in hot-spot agricultural production regions has rarely been investigated. The objectives of this paper are to 1) compare several ML and DL methods, including DNN, 1D-CNN, LSTM, and RF algorithms in estimating winter wheat yield using multi-source data in both county and pixel levels over large areas; 2) understand the major factors that are essential for crop yield estimation; and finally, 3) provide a scalable, simple and inexpensive operational model framework for accurately and timely estimating crop yield.



**Fig. 2.** Visualization of the input data and proposed architectures of DL models. A: the adopted architecture of DNN. B: 1D-CNN. C: LSTM, where A is LSTM cell. Note that 'coord' represents coordinates.

## 2. Materials and methods

### 2.1. Study area

The study area ( $21^{\circ}1'23''\sim42^{\circ}32'43''N$ ,  $92^{\circ}30'15''\sim122^{\circ}59'6''E$ ) includes thirteen provinces, including Shanxi, Shandong, Henan, Jiangsu, Anhui, Hubei, Yunnan, Guizhou, Tianjin, Gansu, Shaanxi, Hebei and Sichuan, one municipality (Beijing), and one autonomous region (Ningxia) in China (Fig. 1). These 15 administrative regions, covering approximately  $4.15 \times 10^6 \text{ km}^2$ , are geographically contiguous. Winters are usually cold and windy, and summers are warm and humid. A total of 629 counties in the regions were analysed, covering the main counties of wheat cultivation in China, with relatively flat topography and fertile soils. Those were selected in the light of having yield data and more than seven planting grids ( $1 \text{ km} \times 1 \text{ km}$ ) for at least 8 years out of a 15-year study period. The winter wheat in China is usually sown at the end of September or the beginning of October and harvested in early or mid-June in the following year (Chen et al., 2018). In general, wheat in those regions is generally irrigated and sufficient fertilization.

### 2.2. Data sources

We collected multi-source data with various spatial and temporal resolutions, including crop planting areas, county-level and field-level yield, climate, satellite, soil properties, and spatial information. An overview of the data was presented in Table S1. Here, we first resampled the gridded data into 1 km, and unified climate and satellite data into monthly time steps. Finally, all input data were aggregated to a mean for each county after being masked by winter wheat planting pixels. All these processes were implemented on the GEE (Google Earth Engine) platform

### 2.3. Crop yield and area

Winter wheat yield data applied in this study included the census yield data at the county level and measured yield data at the field level. The census yields (unit: kg/ha) at the county level for 629 counties from 2001 to 2015 in China were obtained from the Agricultural Yearbook of the provinces and county level statistics bureaus (<http://www.stats.gov.cn>) as described in Tao et al. (2012) and Chen et al. (2018). The outliers were identified and filtered following two rules: (i) they fell outside the range of biophysically attainable yield records. (ii) they fell outside the range of the mean from 2001 to 2015 plus or minus two times of standard deviation (Chen et al., 2018; Zhang et al., 2014). Besides, trial data on wheat cultivars, phenology, yields and management practices from 2011 to 2013 were obtained from China agricultural meteorological experiment stations, which are maintained by China Meteorological Administration (CMA) (<http://data.cma.cn/>). The detail measured yields at the experiment stations allow us to evaluate the accuracy of yield estimation at field scale (Zhang et al., 2014). We only selected the agricultural meteorological experiment stations that had yield records, including 87, 69 and 41 stations in 2011, 2012 and 2013, respectively. The average size of those field in this region is 0.13 ha. Moreover, based on the extracted winter wheat phenology in our previous work (Luo et al., 2020), annual winter wheat planting pixels of China at 1 km resolution were identified from 2000 to 2015 (<https://doi.org/10.6084/m9.figshare.8313530>).

### 2.4. Remote sensing data

The remote sensing data included Terra-MODIS (moderate resolution imaging spectroradiometer) 16-day 1 km EVI (enhanced vegetation index) products (MOD13A2 V6) from 2001 to 2015 and the DEM (digital

elevation model) with  $90 \times 90$  m from Shuttle Radar Topography Mission (SRTM) digital elevation dataset (<http://srtm.csi.cgiar.org>). Compared with the traditional vegetation index-NDVI (normalized difference vegetation index), EVI can better indicate crop canopy foliage and chlorophyll content; therefore, it is closely related to biomass and crop yield (Kang, 2018).

## 2.5. Climatic time series and soil properties

Climate variables, including monthly maximum temperature (Tmmx), minimum temperatures (Tmmn), and precipitation (Pre), were obtained from TerraClimate datasets (<http://doi:10.1038/sdata.2017.191>) in the GEE platform. The monthly datasets of necessary climate variables with a high spatial resolution ( $1/24^\circ$ ,  $\sim 4$  km) for global terrestrial surfaces from 1958 to 2018 were produced by (Abatzoglou et al., 2018) to cover the growing seasons of winter wheat during 2001–2015. In addition to weather data, soil properties including soil depth, soil texture, organic carbon content, pH, cation exchange capacity, and bulk density for the topsoil layer (0–30 cm) and the subsoil layer (30–100 cm) at  $0.00833^\circ$  ( $\sim 1$  km) were also collected and are detailed in Shangguan et al. (2012) (<http://globalechange.bnu.edu.cn>).

The sequential EVI were also resampled to monthly time steps by the MVC (maximum synthesis method). To incorporate spatiotemporal correlations, geographic coordinates and elevation were also considered in the feature space as non-sequential variables. Encoding the geographic coordinates (latitude and longitude) in the feature space accounted for spatial autocorrelation, as nearby counties usually have similar crop yields in a certain year (Kang, 2018; You et al., 2017). Then, all data, including EVI, climate variables, geographic encoding, elevation and soil properties, were masked by winter wheat pixels and aggregated to a mean over 629 counties. Finally, there were four sequential data (EVI, maximum and minimum temperatures, and precipitation) from September to June in the next year and non-sequential features, including soil properties (14 features) and geographic properties (elevation, latitude and longitude).

## 2.6. Methods

DL models are complex non-linear mappings that can learn hierarchical representations of the data. DNN, CNN and LSTM are some typical architectures of DL. To compare three popular DL and ML algorithms in the yield estimation of winter wheat in China, we proposed DNN, 1D-CNN, LSTM architectures and RF models. Detailed information about the adopted network structures and parameter setting are explained in the following sections.

## 2.7. DNN model

A DNN, a special type of feed-forward artificial neural network, is only composed of fully connected layers. All fully connected layers take a vector  $x \in R^n$  as inputs followed by a non-linear function  $f(\cdot)$  (usually a rectified linear unit (ReLU) or tanh) and finally output a vector  $c \in R^n$ , which can be expressed as follows:

$$c = f(Wx + b) \quad (1)$$

The  $W \in R^{n \times n}$  represents the weight matrix, and  $b \in R^n$  is the bias. In this research, five fully connected layers were used to build a DNN (Fig. 2A), which has different neurons in different layers (the first to fifth layers are 528, 384, 128, 64 and 32 neurons, respectively). After trying deeper network structures and more neurons, we found that this structure can achieve the stable and optimal performance in our cases and offer the best balance between prediction accuracy and limited overfitting. All weights were initialized by the Xavier initialization method (Glorot and Bengio, 2010). Batch normalization was used before activation for all layers. The optimizer was RMSprop with a learning rate of 0.1 %, which was divided by 0.1 every 100 iterations. We used a

dropout rate of 0.5 after the hidden layers. Models were run for 2000 maximum iterations with a mini-batch size of 250. A ReLU activation function was used (Goodfellow et al., 2016) for all hidden layers, which did not have any activation function for the last layer. The L2 regularization was used (Khaki and Wang, 2019) for all hidden layers to avoid overfitting.

## 2.8. 1D-CNN model

A CNN, a feed-forward artificial neural network with alternating convolutional and subsampling layers (Goodfellow et al., 2016), is mainly built by fully connected, convolutional and pooling layers. The traditional CNNs are designed for 2D image classification (Kiranyaz et al., 2015). However, 1D-CNNs have recently been successfully used for electrocardiogram beat classification (Kiranyaz et al., 2015), structural damage detection systems (Abdeljaber et al., 2017) and fault detection in high-power engines (Ince et al., 2016), all achieving state-of-the-art performance. Compared with traditional 2D-CNNs, the main difference of 1D-CNNs is the use of 1D arrays instead of 2D matrices for both kernels and feature maps. In 1D-CNNs, the 1D forward propagation from the previous convolution layer,  $l - 1$ , to the input of a neuron in the current layer,  $l$ , can be expressed as follows:

$$x_k^l = b_k^l + \sum_{i=1}^{N_l-1} \text{con1D}(W_{ik}^{l-1}, s_i^{l-1}) \quad (2)$$

$x_k^l$ ,  $b_k^l$  and  $s_i^{l-1}$  are the input data, a scalar bias of the  $k_i^{l-1}$  neuron at layer,  $l$  and the output of the neuron at layer  $l - 1$ , respectively.  $w_{ik}^{l-1}$  represents the kernel weight from the  $i^{\text{th}}$  neuron at layer  $l - 1$  to the  $k^{\text{th}}$  neuron at layer  $l$ .  $y_k^l$ , the intermediate output of the neuron can then be defined as follows:

$$y_k^l = f^*(x_k^l) \quad (3)$$

$$s_k^l = y_k^l \downarrow ss \quad (4)$$

The  $s_k^l$  and  $\downarrow ss$  represent the output of the neuron and the down-sampling operation with  $ss$ , respectively. In this study, we first adopted a novel 1D-CNN architecture that used two separate branches to integrate sequential and non-sequential data (Fig. 2B). In the first branch, learning features from EVI and weather data (maximum and minimum temperatures, and precipitation), were analogous to a scaled-down version of VGGNet, which is one of classic CNN architectures. This branch has one maxpooling layer, one fully connected layer and four convolutional layers. For the second branch, a structure with one max-pooling layer, one fully connected layer and two convolutional layers was used to extract features from the soil and geographic properties (DEM, latitude and longitude). Other parameter (maximum iterations, mini-batch size, ReLU activation function, learning rate and optimizer) settings were the same as for the DNN model (section 3.1).

## 2.9. LSTM model

An LSTM, a basic form of RNN (recursive neural network), needs to input sequential data (Hochreiter and Schmidhuber, 1997). This model maintains a chain structure that simulates the time steps in crop growth modelling. Each time step ( $h_t$ ) depends on the previous step ( $h_{t-1}$ ) and outside input ( $x_t$ ), then produces output ( $o_t$ ) at this moment and offers this time step ( $h_t$ ) for the next step. Finally, a fully connected layer maps the output. In this research, we referenced the architecture by Kang (2018), who used LSTMs for crop yield prediction (Fig. 2C). The sequential data, including EVI and weather data, were dealt with using LSTMs. Finally, a fully connected layer maps the output of the last time step cell to the output node. The LSTM has time steps and three hidden layers, and each LSTM has 128 hidden units. The non-sequential data (DEM, latitude and longitude) are appended to the last hidden state, which is then fully connected to the output layer with yield. Similar to

**Table 1**The RMSE and  $R^2$  for testing data of county-level model performance.

| Year   | $R^2$ |      |      |      | RMSE(kg/ha) |        |        |        |
|--------|-------|------|------|------|-------------|--------|--------|--------|
|        | DNN   | CNN  | LSTM | RF   | DNN         | CNN    | LSTM   | RF     |
| 2011   | 0.84  | 0.87 | 0.87 | 0.89 | 695.63      | 628.00 | 651.85 | 561.18 |
| 2012   | 0.84  | 0.86 | 0.89 | 0.90 | 793.02      | 959.64 | 562.40 | 646.87 |
| 2013   | 0.83  | 0.84 | 0.84 | 0.84 | 842.75      | 729.43 | 826.18 | 878.26 |
| 2014   | 0.85  | 0.88 | 0.87 | 0.90 | 679.08      | 653.28 | 624.71 | 561.93 |
| 2015   | 0.87  | 0.86 | 0.89 | 0.89 | 702.00      | 866.81 | 624.41 | 685.14 |
| median | 0.85  | 0.86 | 0.87 | 0.88 | 742.49      | 767.43 | 657.91 | 666.68 |

the 1D-CNN model, other parameter (maximum iterations, mini-batch size, ReLU activation function, dropout, learning rate and optimizer) settings were the same as those for the DNN model. In this study, the Keras, which is a Python deep learning library, was used to construct DNN, 1D-CNN and LSTM models.

### 2.9.1. RF model

RF, an ensemble-learning method (Breiman, 2001; Cao et al., 2019), is one of the most effective ML techniques and can be used for classification and regression problems. It operates by constructing many decision trees and outputs the predictions by combining decisions from a sequence of base models in the case of regression problems (Liaw and Wiener, 2002). Moreover, the RF model is an additive model; more formally, we can express this model as follows:

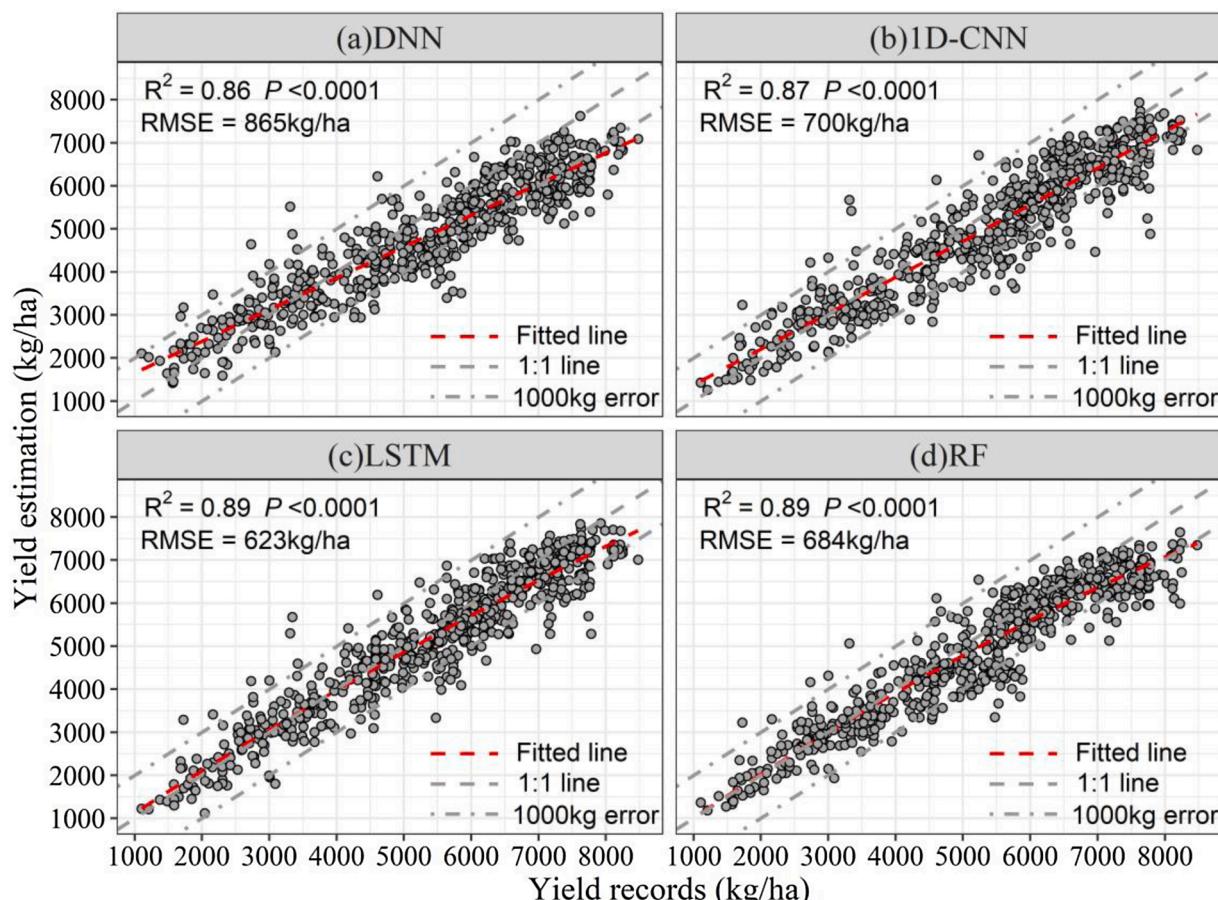
$$g(x) = f(x) + f_1(x) + f_2(x) + \dots + f_n(x) \quad (5)$$

where the model  $g(x)$  is the sum of simple base models  $f_n(x)$ . Each  $f(x)$  is a simple decision tree. All the base models are independently built using a different subsample of the data. Furthermore, the OOB (out-of-bag)

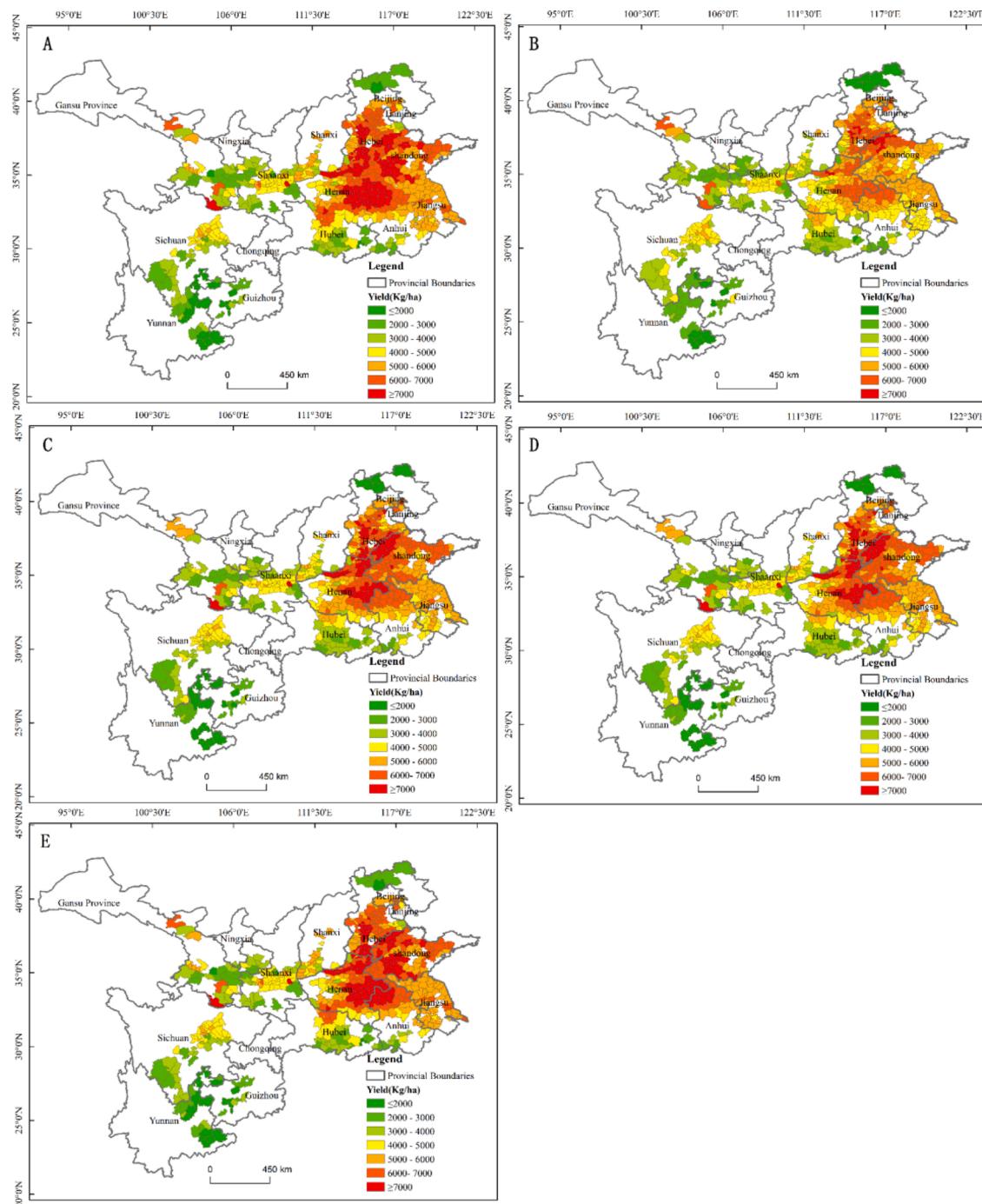
error was used to assess the generalization error rate and the importance of variables. In this study, the scikit-learn, which is an ML library of Python, was used to develop the RF model. After a primary analysis, the number of trees and the tree number of predictive variables used to split the nodes were set to 1500 and 8, respectively, to analyse the joint contributions of subsets of features while maintaining a fast convergence during iterations.

### 2.10. Validation

In this study, we adopted the following training and testing scenarios to evaluate yield estimation accuracy. First, yield estimations were made at ten time months of a year (i.e., from September to June in the next year), which is the main growing season of winter wheat in China. Finally, to assess the practicality of these models, we recursively performed hindcasting for each of the years 2001–2015 to evaluate whether the models can generalize in different years. For example, the data for 2001–June 2014 were collected as training data to predict crop yield in June 2015, and then the data for 2001–2013 and 2015 were applied to



**Fig. 3.** Yield estimation vs. records in county scale for 2015. Result achieved by DNN (A), 1D-CNN (B), LSTM(C) and RF(D).



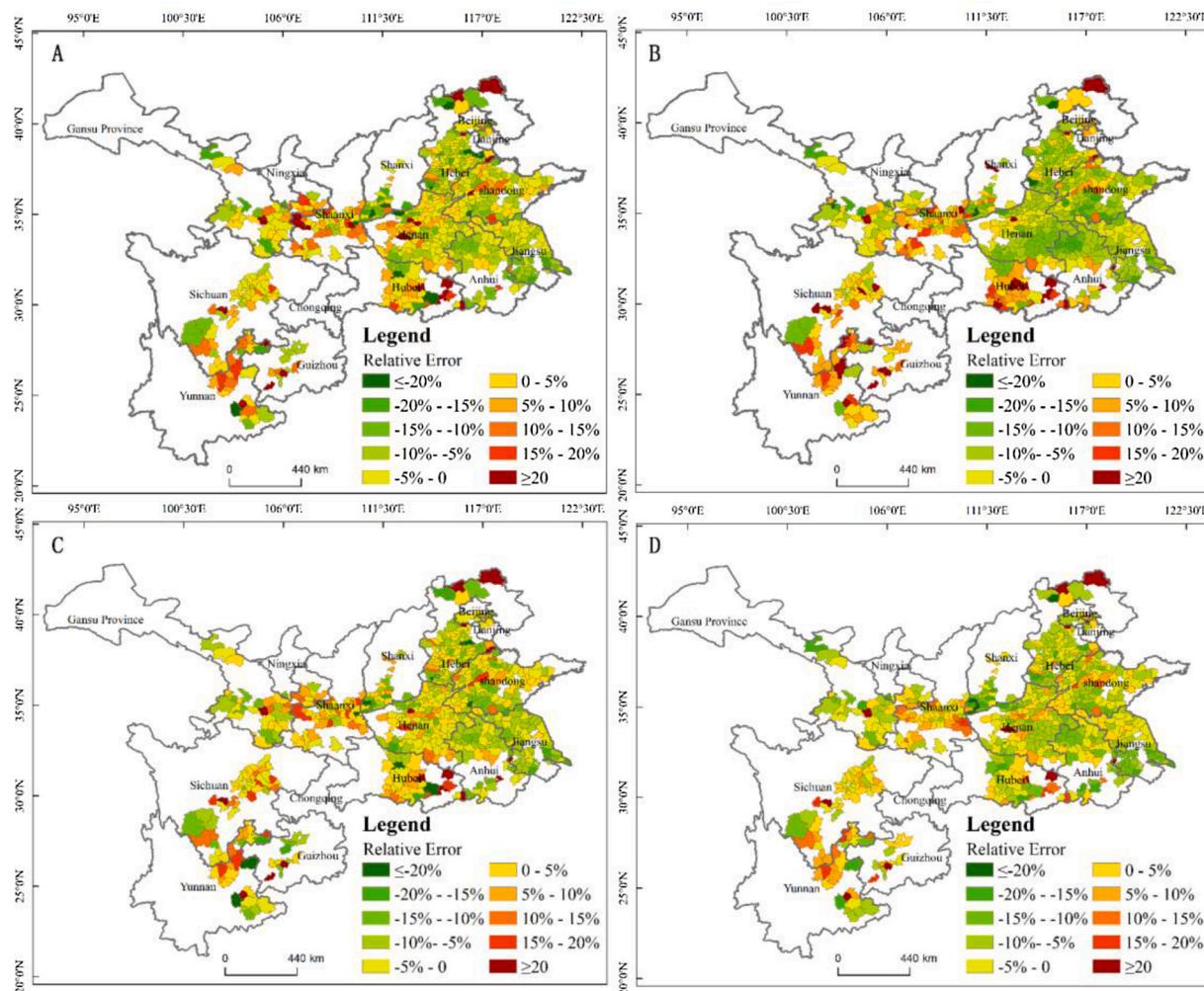
**Fig. 4.** Spatial pattern of yield estimations for 2015 by DNN (A), 1D-CNN(B), LSTM(C), and RF(D), as well as the official census yields at county scale in 2015 (E).

predict crop yield for 2014. Then, the data for 2001–2012 and 2013–2015 were applied to predict crop yield for 2014, and so on, for each of the individual years back to 2011. Certainly, we would never have future data to predict the current data, but these hypothetical scenarios can provide more out-of-sample validation trials to increase the understanding of the model performance (Johnson, 2014). In addition, such a design was also aimed at building a stable, robust and operational methodology for crop yield estimation in the future. We also calculated the RMSE (root mean square error) and  $R^2$  (coefficients of determination) between the validation data and estimates of the trained model (DNN, 1D-CNN, LSTM and RF) to validate yield estimation accuracy.

### 3. Results

#### 3.1. Model comparisons of yield predictions at the county level

The RMSE (kg/ha) and  $R^2$  of yield predictions at the county level by the DNN, 1D-CNN, LSTM, and RF models are shown in Table S2 (training data) and Table 1 (testing data). To account for the random initialization and dropout during DL model training, the results were averaged over 10 runs. Each row corresponds to predictions made for that year, using a model trained on data from all remaining years. RMSE and  $R^2$  were fairly stable during the analysis years for winter wheat at the county scale, although in 2012, the 1D-CNN model had a relatively high RMSE of 959.64 kg/ha (15.8 % of the average yield). Overall, all four models



**Fig. 5.** Mean errors of yield estimation at county level during 2011-2015 by DNN (A), 1D-CNN (B), LSTM(C), and RF(D).

showed a good predictive capability for winter wheat yield at the county level, with an average  $R^2 \geq 0.85$  and RMSE  $\leq 768$  kg/ha from 2011 to 2015, especially for RF (average  $R^2 = 0.88$  and RMSE = 666.68 kg/ha) and LSTM (average  $R^2 = 0.87$  and RMSE = 657.91 kg/ha) models. There was no prediction beyond the outlier for all the models throughout the years in terms of model performance. The scatter plot showed that the estimates and census winter wheat yields at the county scale were close to the 1:1 line, without obvious non-linear or unequal-variance patterns (Fig. 3). Moreover, we also conducted quadratic fit of four model's results (Fig. S3), but those fitted lines can't pass the significance testing. Thus, the three DL models and one ML model showed a consistently high performance in yield prediction. Additionally, LSTM and RF were better than DNN and 1D-CNN in winter wheat yield prediction in China, based on the results from 2011 to 2015.

### 3.2. Spatial patterns of yield predictions at the county level

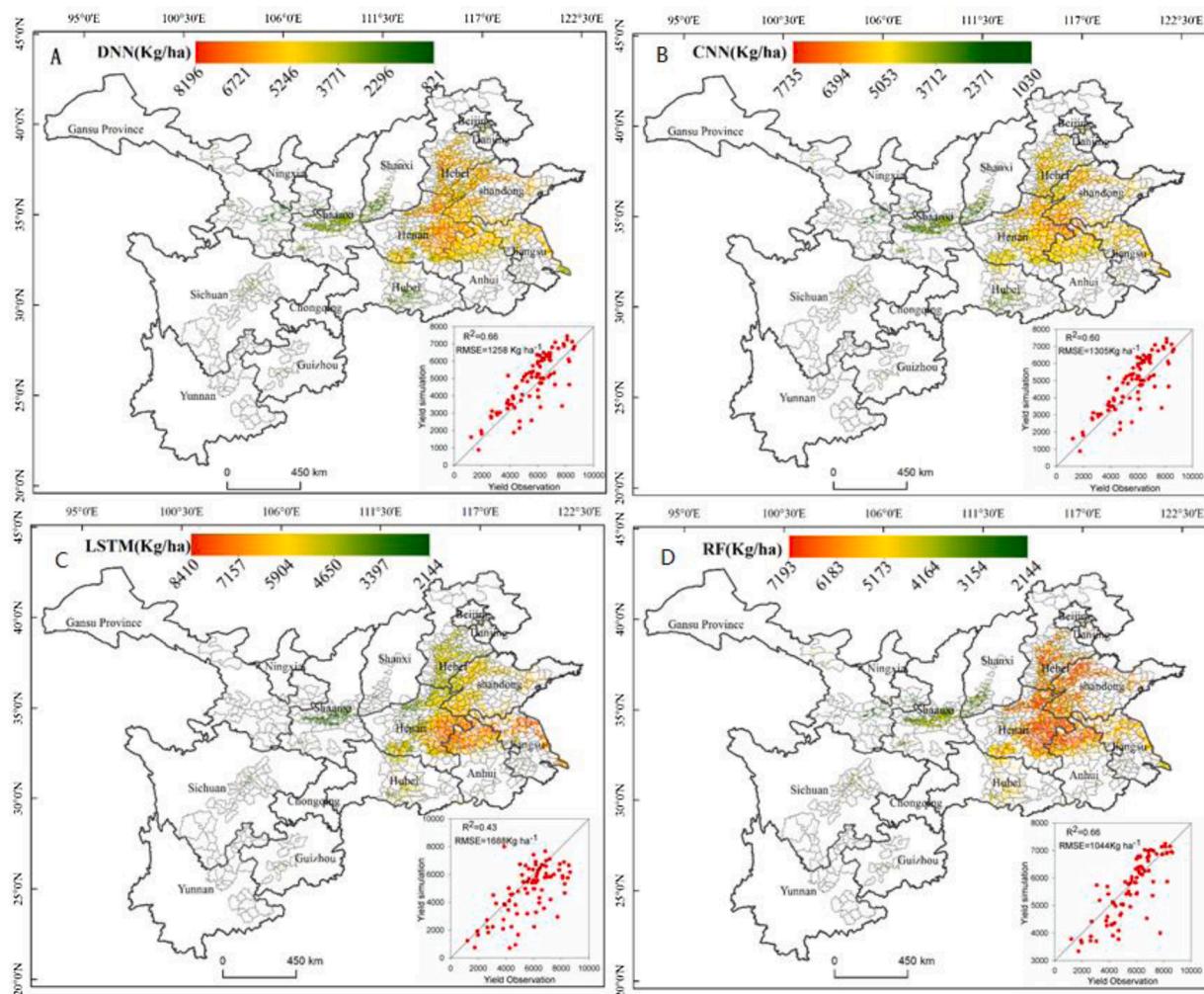
The spatial patterns of yield estimations for 2015 by the four models, which were in good agreement with the official census yields (Fig. 4E), especially for the 1D-CNN, LSTM and RF models, are presented in Fig. 4A-B. The high-yielding counties were mainly located in the northeast of the study area, i.e., the North China Plain. In contrast, the low-yielding counties were sporadically distributed in the southwest of the study area.

The relative error patterns of estimated winter wheat yields in 2015 and mean relative error during 2011–2015 across the entire study area

**Table 2**  
Mean error of four models in main provinces from 2011-2015 (%).

| Province    | DNN   | CNN   | LSTM  | RF    | Avg   |
|-------------|-------|-------|-------|-------|-------|
| Shanxi      | -7.30 | -2.87 | -4.73 | -9.64 | -6.13 |
| Ningxia Hui | -3.82 | -8.80 | -6.22 | -9.03 | -6.97 |
| Shandong    | -2.30 | -5.90 | -2.76 | -3.22 | -3.55 |
| Henan       | -1.86 | -6.40 | -2.08 | -4.21 | -3.64 |
| Jiangsu     | -6.38 | -5.87 | -6.24 | -5.88 | -6.09 |
| Anhui       | -6.39 | -7.18 | -5.56 | -6.86 | -6.50 |
| Hubei       | 4.77  | 10.10 | 4.65  | 0.81  | 5.08  |
| Yunnan      | 4.76  | 10.05 | 1.57  | 2.24  | 4.65  |
| Guizhou     | 8.89  | 12.66 | 2.84  | 4.54  | 7.23  |
| Tianjin     | -3.63 | 3.11  | -1.64 | -2.18 | -1.09 |
| Gansu       | 0.60  | -3.15 | 0.77  | -3.28 | -1.27 |
| Shaanxi     | 4.66  | 4.50  | 3.88  | 2.51  | 3.89  |
| Hebei       | -3.28 | -4.30 | -2.26 | -3.38 | -3.31 |
| Beijing     | -1.82 | -3.65 | -2.79 | -2.95 | -2.80 |
| Sichuan     | 0.51  | 3.47  | 2.38  | 1.28  | 1.91  |

are presented in Fig. S1 and Fig. 5, respectively. The relative error patterns produced by the four models were closely similar in 2015, with the high positive error ( $\geq 20\%$ ) mainly distributed in Sichuan, Hubei, and Shaanxi provinces (Fig. S1). However, the patterns of high negative error ( $\leq -20\%$ ) produced by the four models were different. For example, the DNN and LSTM models produced less negative error, but the 1D-CNN model produced some high negative errors in the middle of the study areas. The 1D-CNN model had the largest error for winter



**Fig. 6.** Winter wheat pixel-level yield maps in 2011 by DNN (A), 1D-CNN (B), LSTM(C), and RF(D), insets represent that estimated yields by DNN (A), 1D-CNN (B), LSTM(C) and RF(D) vs. station yields at field scale for 2011. Note: The insets represent the scatter plot of estimated yields vs. measured yields at field level scale in 2011.

wheat yield estimation in 2015 at the county level. The mean relative error during 2011–2015 showed that the relative error decreased with the number of years (Fig. S1 and Fig. 5). Overall, all four models had relatively small errors during 2011–2015. However, there were few counties with high negative error ( $\leq -20\%$ ), which were sporadically distributed in the study areas. At the province level (Table 2), Ningxia, Anhui, Shanxi, Jiangsu, and Hubei had relatively large errors during 2011–2015 ( $\geq 5\%$ ).

### 3.3. Model comparisons on yield estimations at a field level

To estimate crop yield at a field level, crop yield validation at a field level was performed by creating yield maps at the 1000 m resolution. The estimated crop yields at a field scale in 2011 by the four models are shown in Fig. 6, which had spatial heterogeneity even across a county. The estimated crop yields at the field level were further compared with some observed yield data at 197 agro-meteorological stations across the study areas (Fig. 7). The results showed that there were significant positive correlations for all the combinations of years (2011–2013) and stations, with  $R^2$  ranging from 0.48 to 0.71 and RMSE from 956 kg/ha to 1620 kg/ha (Fig. 7). However, the LSTM model had a large bias at some stations, with an  $R^2$  value of 0.48 and RMSE value of 1620 kg/ha, suggesting that LSTM models were not suitable for yield estimations at a field level, although they had a good performance for yield estimation at a county level (Table 1), with an  $R^2$  value of 0.87 and RMSE value of

651.85 kg/ha. The largest mean  $R^2$  was 0.71 for the DNN model, and the smallest RMSE was 956 kg/ha for the RF model.

### 3.4. Importance of predictor variables in yield estimation

The four types of features in the RF model were used to investigate the importance of these predictor variables. The importance ranking of these features from 2011 to 2015 showed that the four type features had different effects or importance on yield estimation (Fig. 8A). The most important feature in the yield estimation was elevation, followed by latitude, soil properties, and EVI. The importance of all the climate factors (i.e., maximum and minimum temperatures and precipitation) was almost at the bottom of the importance ranking. The effects of EVI, maximum temperature (Tmmx), minimum temperature (Tmnn) and precipitation (Pre) from September to June in the next year on yield estimation are shown in Fig. 8B, which indicates the relative importance of each feature compared to other features in each month. The effects were normalized within each group, i.e., EVI, Tmnn, Tmmx, and precipitation, to make the effects comparable. As shown in Fig. 8B, EVI in April and May had considerable effects on winter wheat yield. In addition, precipitation in December and the temperature from November to January also had relatively larger effects on the yield estimation.

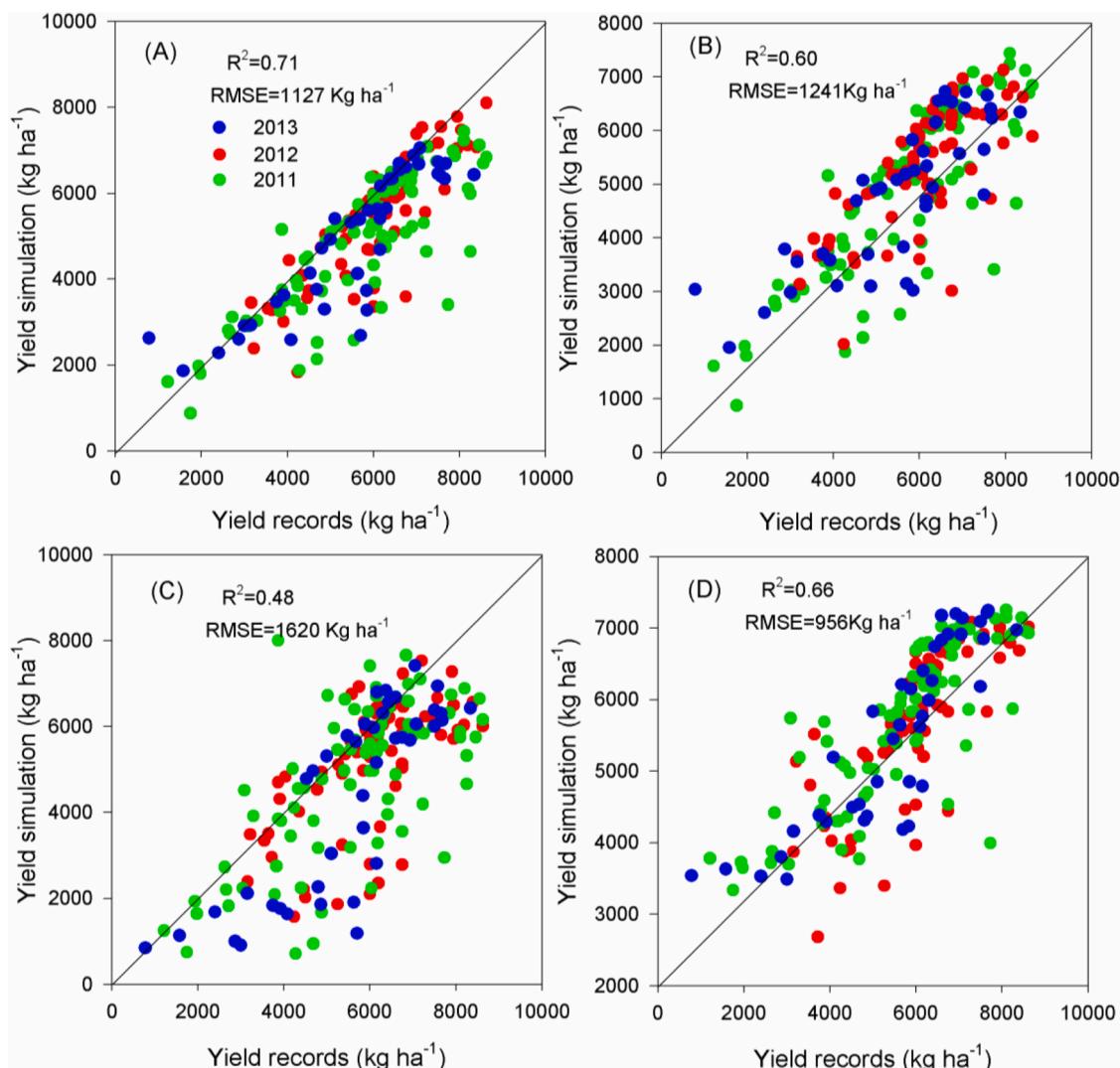


Fig. 7. Scatter plot of yield estimation at field scale by DNN (A), 1D-CNN (B), LSTM(C), and RF(D). vs. measured yields at field scale for 2011-2013.

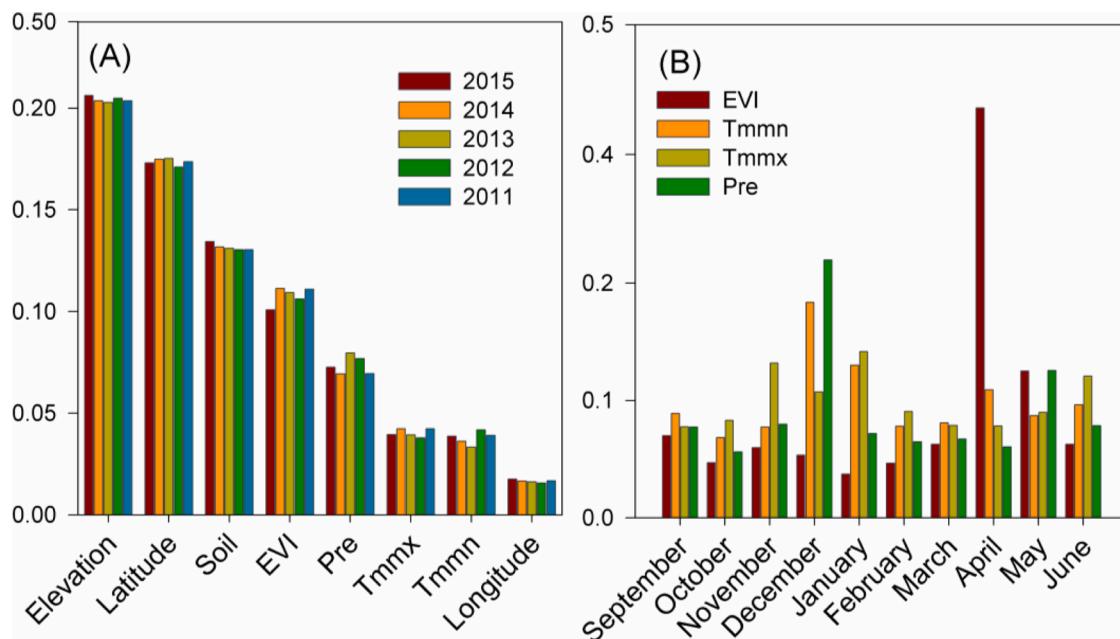
#### 4. Discussion

##### 4.1. DL and ML model performance comparison in yield estimation

To our knowledge, there have been limited studies that estimate crop yield over a large area with DNN, 1D-CNN, LSTM and RF models based on remote sensing data, weather data, soil properties and other spatial data. In this study, weather data, soil properties and spatial data were collected and pre-processed using the GEE platform. All the data are publicly available and can be easily obtained for any country. Then, three typical DL architectures (DNN, 1D-CNN and LSTM) and one ML algorithm (RF) were trained and tested from 2011 to 2015. The results were tested at both the county level and the field level. Recently, DL has been increasingly applied in various domains and has exhibited better performance than traditional techniques (LeCun et al., 2015; Yang et al., 2019), such as image recognition (Krizhevsky et al., 2012; Tompson et al., 2014), speech recognition (Hinton et al., 2012), natural language processing (Collobert et al., 2011) and other applications. However, a review of the basic application of DL algorithms to agriculture documented that there was only one report about yield estimation (Kamilaris and Prenafeta-Boldú, 2018). Although subsequent progress in the field of DL for yield estimation has been noticed, applying DL to agriculture is still nascent (Koirala et al., 2019), especially in crop yield estimation. Most of the studies still focus more on computer vision and image

analysis in agricultural applications (Kamilaris and Prenafeta-Boldú, 2018).

In this study, we found that overall, all four models exhibited good performance for winter wheat yield estimation at a county level from 2011 to 2015. However, only the DNN and RF models had good performance at the field level. The model comparisons showed that the performance of ML was not always worse than DL at both the county and field levels. It seemed that the RF model performed better than the other DL models to some extent. We speculate there are two main reasons. First, the training of the DL model requires a large number of datasets (Coltin et al., 2016; Reyes et al., 2015), but we only had 8806 and 629 samples as training data and test data each time, respectively. Second, one of the advantages of DL is that it can automatically extract new features from input data and reduce the need for feature engineering (Kamilaris and Prenafeta-Boldú, 2018). While in this present study, many features, including remote sensing data, weather, soil information, and geographic properties, have been selected by users, the potential of DL may not be sufficiently utilized. These selected features have strong correlations with crop yields and can be used to estimate crop yields even with traditional ML models. Overall, the three DL and RF models achieved better results at a county level than some previous studies in winter wheat yield estimation (Chen et al., 2018; Han et al., 2020). For example, Chen et al. (2018) shows that crop model assimilated by crop phenology and leaf area index (LAI) can only achieve  $R^2 = 0.42$  and



**Fig. 8.** Importance ranking of the features in yield estimation from 2011-2015(A) and bar plot of estimated effects of four sequential data for 10 months and soil properties, starting from September to June in next year(B). The vertical axes were normalized across all feature components to make the effects comparable. The Tmmx, Tmmn, and Pre refer to monthly maximum temperatures ( $^{\circ}\text{C}$ ), minimum temperatures ( $^{\circ}\text{C}$ ), and precipitation (mm), respectively.

RMSE = 737 kg/ha in North China Plain. Han et al. (2020) indicates their yield models can achieve the highest  $R^2 = 0.75$  and the lowest RMSE = 923.07 kg/ha in same region. all four models are acceptable for yield estimation in the study area at the county level. Nevertheless, the three DL models require a long training time and considerable computer power.

#### 4.2. Importance of features in yield estimations

The results showed that the most important feature in the yield estimation was elevation, followed by longitude, soil properties, and EVI. The importance of all the climate factors (i.e., maximum and minimum temperatures and precipitation) was almost at the bottom of the importance ranking. The results are supported by some previous studies (Liu et al., 2019). The underlying reasons may be that elevation, latitude, and soil properties represent comprehensive features or information of a county or a field over a long time, while climate factors only represent a part of the information related to crop yield for a specific period. High-yielding areas can be characterized by fertile soils, water conditions, well-educated farmers, good technologies, well-equipped irrigation facilities, and suitable climate conditions. All these features can be comprehensively represented by spatial location, but not by EVI or climate factors, which are in-line with the existing prior work, for example, Wang et al. (2019) applies spatial analysis to prove that the spatial factors (e.g. elevation) not only controls the distribution of temperature, sunlight and soil, but effects phenological period of crop. Those consequently directly impact crop growth. Furthermore, we also prove that the spatial location information (i.e. elevation and latitude) have significant correlations between climate factors (e.g. temperature and precipitation), with the exception of elevation and precipitation ( $p < 0.0001$ ) (Fig. S3). However, climate factors may be more important in affecting the inter-annual variability of crop yield (Ceglar et al., 2016). The results also showed that the vegetation index (EVI) in April and May had considerable effects on winter wheat yield, and precipitation in December and temperature from November to January had relatively larger effects on the yield estimation. These are also supported by some previous studies (Feng et al., 2010; Ren et al., 2010). The underlying reasons may be that EVIs in April and May are significantly correlated

with the yield of winter wheat at the heading stage (Ren et al., 2010). The temperature from November to January is an important predictor for winter wheat in the regions because it is closely related to cold and drought stress on wheat (Tao et al., 2017)

#### 4.3. Implications for improving yield estimation with DL

The results showed that both DL and ML models performed well in crop yield estimation at a county level. The results suggest that the framework has the potential to provide a scalable, simple and inexpensive method for timely and accurately predicting crop yields using publicly available data. In this study, all features were aggregated to a monthly scale, and although performance was good at the county level, all four models had relatively poor performance at a pixel level, especially for the LSTM model. We speculate that monthly EVI and weather data cannot accurately reflect crop growth and development. The EVI at the 8-day period or 16-day period might better incorporate crop growth and weather information (Kang, 2018). In addition, both DL and ML models are black boxes. It is difficult to produce testable hypotheses that could potentially provide biological insights because of their complex model structure. Combining crop models and DL/ML models for yield estimation, forecasts, and disaster monitoring in large regions is recommended.

#### 5. Conclusions

In this study, we first processed all data (including climate variables, satellite, DEM, soil properties and coordinates data) on the GEE platform and then used the four models, including three DL models (DNN, 1D-CNN and LSTM) and one ML model (RF) to predict winter wheat yield from 2011 to 2015. Finally, we compared accuracy and generalization for all years at both the county and field levels. Overall, all four models performed well at the county level, but only the DNN and RF models performed well at the field level. Our findings demonstrated a new scalable, simple and inexpensive framework for estimating winter wheat yield on a regional scale with publicly available data and GEE platforms, which can potentially be applied to areas with sparsely observed data and worldwide for estimating crop yield. The framework can be further

improved by combining crop models and more tailored EVI data for crop yield estimation, forecasts, and disaster monitoring in large regions.

## CRediT authorship contribution statement

**Juan Cao:** Methodology, Software, Formal analysis, Investigation, Writing - original draft. **Zhao Zhang:** Conceptualization, Validation, Visualization, Funding acquisition. **Yuchuan Luo:** Validation, Formal analysis, Visualization. **Liangliang Zhang:** Resources, Writing - review & editing, Data curation. **Jing Zhang:** Resources, Writing - review & editing, Data curation. **Ziyue Li:** Writing - review & editing. **Fulu Tao:** Writing - review & editing.

## 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 Basic Research Program of China (41977405, 31761143006).

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.eja.2020.126204>.

## 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.
- Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., Inman, D.J., 2017. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* 388, 154–170.
- 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.
- Battisti, R., Sentelhas, P.C., Boote, K.J., 2017. Inter-comparison of performance of soybean crop simulation models and their ensemble in southern Brazil. *Field Crops Res.* 200, 28–37.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* 210, 35–47.
- Cao, J., Zhang, Z., Wang, C., Liu, J., Zhang, L., 2019. Susceptibility assessment of landslides triggered by earthquakes in the Western Sichuan Plateau. *Catena* 175, 63–76.
- Ceglar, A., Toreti, A., Lecerf, R., Van der Velde, M., Dentener, F., 2016. Impact of meteorological drivers on regional inter-annual crop yield variability in France. *Agric. For. Meteorol.* 216, 58–67.
- Chen, Y., Zhang, Z., Tao, F., Wang, P., Wei, X., 2017. Spatio-temporal patterns of winter wheat yield potential and yield gap during the past three decades in North China. *Field Crops Res.* 206, 11–20.
- 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.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P., 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* 12 (Aug), 2493–2537.
- Coltin, B., McMichael, S., Smith, T., Fong, T., 2016. Automatic boosted flood mapping from satellite data. *Int. J. Remote Sens.* 37 (5), 993–1015.
- Dodds, F., Bartram, J., 2016. The Water, Food, Energy and Climate Nexus: challenges and an Agenda for Action. Routledge.
- Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C., Moore 3rd, B., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens. Environ.* 185, 142–154.
- FAO, 2015. The State of Food Insecurity in the World. Meeting the 2015 International Hunger Targets: Taking Stock of Uneven Progress.
- Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics 249–256.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT press.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27.
- Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D. B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sens. Environ.* 199, 333–349.
- Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R., 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342 (6160), 850–853.
- He, Z., Xia, X., Zhang, Y., 2010. Breeding noodle wheat in China. *Asian Noodl.: Sci. Technol. Process.* 1–23.
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process. Mag.* 29.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., Wu, W., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121.
- Ince, T., Kiranyaz, S., Eren, L., Askar, M., Gabbouj, M., 2016. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* 63 (11), 7067–7075.
- 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.
- Joshi, A.R., Dinerstein, E., Wikramanayake, E., Anderson, M.L., Olson, D., Jones, B.S., Seidensticker, J., Lumpkin, S., Hansen, M.C., Sizer, N.C., 2016. Tracking changes and preventing loss in critical tiger habitat. *Sci. Adv.* 2 (4), e1501675.
- Kamilaris, A., Prenafta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. *Comput. Electron. Agric.* 147, 70–90.
- Kang, Y., 2018. County-Level Corn Yield Forecast With Deep Learning and Satellite Data. Available online: <http://homepages.cae.wisc.edu/> (accessed on 04 June 2019).
- Khaki, S., Wang, L., 2019. Crop yield prediction using deep neural networks. *Front. Plant Sci.* 10, 621.
- Kiranyaz, S., Ince, T., Gabbouj, M., 2015. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* 63 (3), 664–675.
- Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019. Deep learning – method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* 162, 219–234.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems.
- Kuwata, K., Shibasaki, R., 2015. Estimating crop yields with deep learning and remotely sensed data. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 858–861.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444.
- Liau, A., Wiener, M., 2002. Classification and regression by randomForest. *R news* 2 (3), 18–22.
- Liu, J., He, X., Wang, P., Huang, J., 2019. Early prediction of winter wheat yield with long time series meteorological data and random forest method. *Transact. CSAE* 35 (6).
- Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Res.* 143, 56–64.
- 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.
- Luo, Y., Zhang, Z., Chen, Y., Li, Z., Tao, F., 2020. ChinaCropPhen1km: a high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on LAI products. *Earth Syst. Sci. Data* 12 (1), 197–214.
- Macdonald, R.B., Hall, F.G., 1980. Global crop forecasting. *Science* 208 (4445), 670–679.
- Pekel, J.F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540 (7633), 418–422.
- 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.
- Reyes, A.K., Caicedo, J.C., Camargo, J.E., 2015. Fine-tuning Deep Convolutional Networks for Plant Recognition. CLEF (Working Notes), p. 1391.
- Russello, H., 2018. Convolutional Neural Networks for Crop Yield Prediction Using Satellite Images. Master's Thesis. University of Amsterdam, Amsterdam, The Netherlands.
- Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sens. Environ.* 131, 215–231.
- Shangguan, W., Dai, Y., Liu, B., Ye, A., Yuan, H., 2012. A soil particle-size distribution dataset for regional land and climate modelling in China. *Geoderma* 171–172, 85–91.
- Singha, M., Dong, J., Zhang, G., Xiao, X., 2019. High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Sci. Data* 6 (1), 26.
- Sturrock, H.J., Cohen, J.M., Keil, P., Tatem, A.J., Le Menach, A., Ntshalintshali, N.E., Hsiang, M.S., Gosling, R.D., 2014. Fine-scale malaria risk mapping from routine aggregated case data. *Malar. J.* 13 (1), 421.
- Tao, F., Zhang, Z., Zhang, S., Zhu, Z., Shi, W., 2012. Response of crop yields to climate trends since 1980 in China. *Clim. Res.* 54 (3), 233–247.

- Tao, F., Xiao, D., Zhang, S., Zhang, Z., Rötter, R.P., 2017. Wheat yield benefited from increases in minimum temperature in the Huang-Huai-Hai Plain of China in the past three decades. *Agric. For. Meteorol.* 239, 1–14.
- Tompson, J.J., Jain, A., LeCun, Y., Bregler, C., 2014. Joint training of a convolutional network and a graphical model for human pose estimation. *Adv. Neural Inf. Process. Syst.* 1799–1807.
- Wang, C., Zhang, Z., Zhang, J., Tao, F., Chen, Y., Ding, H., 2019. The effect of terrain factors on rice production: a case study in Hunan Province. *J. Geogr. Sci.* 29, 287–305.
- Yang, Q., Shi, L., Han, J., Zha, Y., Zhu, P., 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Res.* 235, 142–153.
- You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep gaussian process for crop yield prediction based on remote sensing data. *Thirty-First AAAI Conference on Artificial Intelligence*.
- Zhang, T., Yang, X., Wang, H., Li, Y., Ye, Q., 2014. Climatic and technological ceilings for Chinese rice stagnation based on yield gaps and yield trend pattern analysis. *Glob. Chang. Biol.* 20, 1289–1298. <https://doi.org/10.1111/gcb.12428>.